PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 :
C12N 15/57, 9/64, C07K 16/40, C12Q 1/37, A61K 35/00, 38/48

A2

(11) International Publication Number: WO 98/54333

(43) International Publication Date: 3 December 1998 (03.12.98)

(21) International Application Number: PCT/US98/11415

(22) International Filing Date: 2 June 1998 (02.06.98)

(30) Priority Data:

60/047,369	2 June 1997 (02.06.97)	US
60/052,389	14 July 1997 (14.07.97)	US

(63) Related by Continuation (CON) or Continuation-in-Part (CIP) to Earlier Applications

US	60/047,369 (CIP)
Filed on	2 June 1997 (02.06.97)
US	60/052,389 (CIP)
Filed on	14 July 1997 (14.07.97)

(71) Applicant *(for all designated States except US)*: ACACIA BIOSCIENCES INC. [US/US]; 4136 Lakeside Drive, Richmond, CA 94806 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: ASHBY, Matthew, N. [US/US]; 91 Longfellow Road, Mill Valley, CA 94941 (US). DIMSTER-DENK, Dago, G. [US/US]; 904 Sir Francis Drake Boulevard #8, Kentfield, CA 94904 (US). PHILIPS, John, W. [US/US]; 10 Marinita Avenue, San Rafael, CA 94901 (US).

(74) Agents: HALEY, James, F., Jr. et al.; Fish & Neave, 1251 Avenue of the Americas, New York, NY 10020 (US).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, GW, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

Published

Without international search report and to be republished upon receipt of that report.

With an indication in relation to deposited biological material furnished under Rule 13bis separately from the description.

(54) Title: MAMMALIAN CAAX PROCESSING ENZYMES

(57) Abstract

The present invention provides mammalian DNA sequences that display a high degree of sequence identity to their *Saccharomyces cerevisae* counterparts, *RCE1 and AFC1*. Specifically, complementary DNA (cDNA) sequences of the human and mouse *RCE* functional homologs are provide. Human cDNA sequences encoding proteins having a high degree of amino acid sequence identity to the yeast Afc1p protein are also provided. This invention is also directed to recombinant DNA molecules comprising the mammalian DNA sequences, DNA molecules and antisense RNA molecules which hybridize under stringent hybridization conditions to those DNA sequences, hosts transformed with their recombinant DNA molecules and protein expression products produced by culturing the transformed hosts. Antibodies directed against the protein expression products are also provided. Also provided are assays to identify inhibitors of one or more mammalian CAAX processing enzymes and kits for making the above products and performing the above assays. Finally, this invention provides pharmaceutical compositions comprising an inhibitor of a mammalian CAAX processing enzyme, and methods for treating a CAAX-protein mediated disease or disorder in a patient by administering such a pharmaceutical composition..

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	**ES**	Spain	**LS**	Lesotho	**SI**	Slovenia
AM	Armenia	**FI**	Finland	**LT**	Lithuania	**SK**	Slovakia
AT	Austria	**FR**	France	**LU**	Luxembourg	**SN**	Senegal
AU	Australia	**GA**	Gabon	**LV**	Latvia	**SZ**	Swaziland
AZ	Azerbaijan	**GB**	United Kingdom	**MC**	Monaco	**TD**	Chad
BA	Bosnia and Herzegovina	**GE**	Georgia	**MD**	Republic of Moldova	**TG**	Togo
BB	Barbados	**GH**	Ghana	**MG**	Madagascar	**TJ**	Tajikistan
BE	Belgium	**GN**	Guinea	**MK**	The former Yugoslav Republic of Macedonia	**TM**	Turkmenistan
BF	Burkina Faso	**GR**	Greece			**TR**	Turkey
BG	Bulgaria	**HU**	Hungary	**ML**	Mali	**TT**	Trinidad and Tobago
BJ	Benin	**IE**	Ireland	**MN**	Mongolia	**UA**	Ukraine
BR	Brazil	**IL**	Israel	**MR**	Mauritania	**UG**	Uganda
BY	Belarus	**IS**	Iceland	**MW**	Malawi	**US**	United States of America
CA	Canada	**IT**	Italy	**MX**	Mexico	**UZ**	Uzbekistan
CF	Central African Republic	**JP**	Japan	**NE**	Niger	**VN**	Viet Nam
CG	Congo	**KE**	Kenya	**NL**	Netherlands	**YU**	Yugoslavia
CH	Switzerland	**KG**	Kyrgyzstan	**NO**	Norway	**ZW**	Zimbabwe
CI	Côte d'Ivoire	**KP**	Democratic People's Republic of Korea	**NZ**	New Zealand		
CM	Cameroon			**PL**	Poland		
CN	China	**KR**	Republic of Korea	**PT**	Portugal		
CU	Cuba	**KZ**	Kazakstan	**RO**	Romania		
CZ	Czech Republic	**LC**	Saint Lucia	**RU**	Russian Federation		
DE	Germany	**LI**	Liechtenstein	**SD**	Sudan		
DK	Denmark	**LK**	Sri Lanka	**SE**	Sweden		
EE	Estonia	**LR**	Liberia	**SG**	Singapore		

MAMMALIAN CAAX PROCESSING ENZYMES

BACKGROUND OF THE INVENTION

The eukaryotic cell is made up of a wide variety of specialized structures and membrane-enclosed compartments which create special intracellular environments. Many cellular components, such as proteins, must be targeted to particular intracellular locations for proper function. The lipid bilayer membrane that surrounds the cell is one such specialized cellular environment. Cellular components such as proteins are delivered to and must associate in a particular way with the correct membrane surface of the lipid bilayer. A number of mechanisms which the cell has evolved for targeting proteins to and inserting proteins into the lipid bilayer membrane of the cell have now been described.

Proteins targeted to the lipid bilayer membrane may be situated within the lipid bilayer membrane (i.e., a portion of the protein is inserted into and across the lipid bilayer) and are typically referred to as "integral membrane proteins". Integral membrane proteins possess intrinsic hydrophobic regions (comprising amino acid regions of relatively high intrinsic or apparent hydrophobicity) which are often inserted into the lipid bilayer as the protein is being synthesized.

Alternatively, proteins targeted to the lipid bilayer membrane may be situated at and associate with the membrane surface without being inserted across the membrane. Such proteins are typically referred to as "peripheral membrane proteins".

Peripheral membrane proteins are usually less tightly associated with the lipid bilayer than integral membrane proteins. Peripheral membrane proteins may associate with the lipid bilayer by protein-protein interactions, by hydrophobic interactions between the bilayer and one or more intrinsically hydrophobic regions of the membrane protein, or by the co- or post-translational addition of a lipid group to the membrane protein. The most recently discovered lipid modification involves the covalent attachment of cholesterol biosynthetic intermediates to proteins, and is referred to generically as protein prenylation. See, e.g., W.R. Schafer and J. Rine, Annu. Rev. Genet., 26, p. 209 (1992); S. Clarke, Annu. Rev. Biochem., 61, p. 355 (1992); each of which is incorporated herein by reference.

One class of peripheral membrane proteins processed by the latter mechanism has a terpenoid lipid added to the C-terminal end for such purposes. Proteins belonging to this class (often referred to as prenylated proteins) that possess a C-terminal tetrapeptide CAAX motif (where C is cysteine, A is usually aliphatic, and X can be a variety of different residues) undergo a series of post-translational processing reactions. These reactions include the attachment of either a 15-carbon farnesyl or a 20-carbon geranylgeranyl lipid moiety to the cysteine residue of the CAAX motif (prenylation), proteolytic removal of the C-terminal three amino acids (AAX) and methylesterification of the α-carboxyl group of the terminal prenylated cysteine residue.

Prenylation of CAAX protein substrates is catalyzes by one of two different type I protein prenyl-

transferases. If the X position is occupied by leucine or phenylalanine, the cysteine residue of the protein is geranylgeranylated. If the X position is occupied by any other amino acid, the cysteine residue of the protein is farnesylated. See V. L. Boyartchuk et al., Science, 275, pp. 1796-1800 (1997) and M.N. Ashby, Current Opinion in Lipidol., 9, pp. 99-102 (1998) (incorporated herein by reference).

Prenylated proteins include, for example, many proteins involved in signal transduction such as the Ras oncogenes, fungal pheromones and the γ subunits of trimeric guanosine triphosphate (GTP)-binding proteins (G proteins). They also include a wide range of small GTP-binding proteins that participate in vesicular trafficking. See, e.g, P. J. Casey, Curr. Opin. Cell. Biol., 6, p. 219 (1994); J.E. Rothman and F. T. Wieland, Science, 272, p. 227 (1996). Prenylation is critical for the activity of many of these proteins.

The yeast *Saccharomyces cerevisiae* possesses numerous CAAX-containing proteins including, but not limited to, Ras1p, Ras2p, Mfa1p, Mfa2p, and Ste18p. Yeast have two genes encoding prenylation-dependent CAAX endoproteinases, *AFC1* and *RCE1* (Boyartchuk et al., supra). The Afc1 protein (Afc1p; encoded by *AFC1*) is a zinc protease that participates in the processing of the yeast mating pheromone a-factor. The Rce1 protein (Rce1p; encoded by *RCE1*) contributes to the processing of a-factor and the yeast Ras protein (Ras2p). Rce1p is only a minor contributor in vivo to a-factor processing but appears to be required for Ras2p processing. Yeast have a single gene, *STE14*, which encodes the membrane-associated, CAAX C-terminal methyltransferase (Ste14p) (Ashby et al., Yeast, 9, p. 907 (1993) (incorporated herein by reference)).

Farnesylation of the Ras protein of yeast and

humans is required for proper membrane localization. The Ras protein in normal cells contributes to the orderly cell divisions needed during growth and development. Ras proteins are attached to the inner surface of the lipid bilayer membrane that surrounds each cell. Ras proteins function as a primary intracellular signaling mechanism for cell proliferation. They respond to external stimuli such as growth factors and hormones which bind to the cell membrane and trigger an intracellular cascade of chemical reactions known generally as "signal transduction". These chemical signals lead to the cell nucleus, where they activate gene transcription and the process of cell division. When the extracellular stimulus stops, the internal signaling pathway is deactivated and the cell returns to its resting state.

When dominant mutations occur in the Ras gene (turning it into a Ras oncogene), mutant proteins are produced that lock the signaling pathway in an active state, resulting in uncontrolled cell division. Dominant mutations in the Ras gene are presently the second most common type of mutations in human cancers. They are found in more than 50% of all human cancers including pancreatic cancer (95%), colon cancer (50%), lung cancers (45%) and leukemias (30%).

The functional activity of Ras protein depends on its attachment to the cell membrane, which in turn is mediated by protein prenylation. Farnesylation of the Ras protein in yeast and humans is required for its oncogenic transforming activity. Because the Ras protein is modified sequentially by the three different processing enzymes described above (which farnesylate, cleave and methylate the C-terminal CAAX motif), each of these processing enzymes represents a potential target that could be attacked pharmacologically to block the oncogenic action of mutant Ras protein.

Protein farnesyl transferases have become an

active target for drug development. Inhibitors of protein farnesyl transferases are being developed, for example, as potential therapeutics for Ras-based tumors. See, e.g., F. Tamanoi, Trends Biochem. Sci., 18, p. 349 (1993); N. E. Kohl et al., J. Cell. Biochem. Suppl., 22, p. 145 (1995). The methyltransferase enzyme will also be an attractive target for drug development.

The mammalian counterparts to the yeast CAAX endoproteinase and methyltransferase genes have not been isolated. The mammalian homologs to each of the three yeast CAAX processing enzymes each represent a potential target to block the oncogenic action of mutant Ras protein in tumors, or more generally, to modulate the activity of prenylated peripheral membrane proteins. It would thus be desirable to identify and to isolate the mammalian homologs to the yeast CAAX endoproteinase and methyltransferase genes.

SUMMARY OF THE INVENTION

The present invention provides mammalian DNA sequences that display a high degree of sequence identity to their counterpart genes in *Saccharomyces cerevisiae*, *RCE1* and *AFC1*, encoding prenylation-dependent CAAX endoproteinases. Specifically, complementary DNA (cDNA) sequences of the human and mouse *RCE1* functional homologs are provided. Human cDNA sequences encoding proteins having a high degree of amino acid sequence identity to the yeast Afc1p protein are also provided. This invention is also directed to recombinant DNA molecules comprising the mammalian DNA sequences, DNA molecules and antisense RNA molecules which hybridize under stringent hybridization conditions to those DNA sequences, hosts transformed with the recombinant DNA molecules and protein expression products produced by culturing the transformed

hosts. Antibodies directed against the protein expression products are also provided. Also provided are assays to identify inhibitors of one or more mammalian CAAX processing enzymes and kits for making the above products and performing the above assays. Finally, this invention provides pharmaceutical compositions comprising an inhibitor of a mammalian CAAX processing enzyme, and methods for treating a CAAX-protein mediated disease or disorder in a patient by administering such a pharmaceutical composition.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1A. Amino acid sequence alignment of yeast *RCE* with a human EST.

Figure 1B. Amino acid sequence alignment of yeast AFC1 (NAG2) with a human EST.

Figure 1C. Amino acid sequence alignment of yeast STE14 with a mouse EST.

Figure 2A. Human *hRCE* cDNA (SEQ ID NO:1) and predicted protein sequences (SEQ ID NO:2).

Figure 2B. Mouse *mRCE* cDNA (SEQ ID NO:3) and predicted protein sequences from the first in-frame ATG codon (SEQ ID NO:4).

Figure 2C. Mouse *mRCE* cDNA sequence with additional 5' genomic sequence relative to Figure 2B (SEQ ID NO:5). The first in-frame ATG codon is underlined. The sequence obtained exclusively from the mouse genomic clones is represented from nucleotides 1-152 (SEQ ID NO:5). See (SEQ ID NO:6) for the predicted protein sequence.

Figure 3. Amino acid sequence alignment of the yeast Rce1p protein (RCE1.PRO), the human hRCE protein (HRCE1.PRO) and the mouse mRCE (MRCEORF.PRO) protein. The alignment utilized the Clustal method with PAM250 residue weight table.

Figure 4. Northern blots of *hRCE* mRNA expression levels in various human tissues. The source tissue for the poly(A) RNA is indicated along the top, and the size standards (kbp) along the left side of each blot. The *hRCE* message appears to be approximately 1.4 kbp and was expressed in all tissues examined.

Figure 5. Primer extension analysis of *hRCE* mRNAs. Human placental and pancreatic poly(A) RNA was hybridized to a *hRCE* specific primer and extended using reverse transcriptase. The extension products were resolved by polyacrylamide gel electrophoresis, along with the products of a sequencing reaction to serve as size standards. The transcriptional start sites appear to be in the region from -10 to -26 (relative to the valine codon GTG at position 116-118; Figure 2A).

Figure 6. Yeast halo assays of wild-type strain (W303) (MAT-a; leu2-3,112; ade2-101; his3-11; trp1-1; ura3-1), a strain lacking the *AFC1* gene (Δafc1), a strain lacking the *RCE1* gene (Δrce1), and a strain lacking both *AFC1* and *RCE1* (ΔΔ) are shown in the top four panels. Halo assays of a strain lacking both *AFC1* and *RCE1* genes expressing either no human protein (ΔΔpJP1), *hRCE* expressed from the yeast MEV1 promoter (ΔΔpACA7001), or *hRCE* expressed from the yeast ERG12 promoter (ΔΔpACA7002) are shown in the lower three panels. Halos are present when *hRCE* is expressed in yeast indicating that the human protein is

functional in yeast (see **Example 3**).

Figure 7. Human *hAFC* cDNA and protein sequences.

Figure 8. Amino acid sequence alignment between the full length yeast Afc1p protein (YAFC1.PRO) and the human hAFC protein (HAFC1.PRO). The alignment utilized the Clustal method with PAM250 residue weight table.

Figure 8. Amino acid alignment between the full length yeast Afc1p protein (YAFC1.PRO) and the human hAFC protein (HAFC1.PRO). The alignment utilized the Clustal method with PAM250 residue weight table.

Figure 9. Amino acid alignment between a region of the human hRCE1p protein (SEQ ID NO:2, amino acid residues 166-233) comprising the histidine-rich sequence, HxHH (SEQ ID NO:2, amino acid residues 172-175) and nucleic acid sequences in GenBank. The alignment utilized the Clustal method with PAM250 residue weight table.

Figure 10. Yeast halo assays performed with the following strains: wild-type strain (W303)(MAT-a; leu2-3,112; ade2-101; his3-11; trp1-1; ura3-1); a strain lacking the *RCE1* gene (Δrce1); a strain lacking both *AFC1* and *RCE1* (Δafc1, Δrce1); a strain transformed with a URA3$^+$ marked, high-copy (2μ) plasmid carrying wildtype *hRCE* expressed from the yeast ERG12 promoter (pAB13 wild-type); a strain transformed with a pAB13-derived plasmid in which the histidine at position 172 of *hRCE* was changed to an alanine (pAB15 H172A); a strain transformed with a pAB13-derived plasmid in which the histidine at position 174 of *hRCE* was changed to an alanine (pAB16 H174A); a strain transformed with a pAB13-derived plasmid in which the

histidine at position 175 of *hRCE* was changed to an alanine (pAB17 H175A). Halos are present when *hRCE* is expressed in yeast indicating that the human protein is functional in yeast (see **Example 3**).

Figure 11. Heat shock assays in wild-type yeast cells (W303) (*MATa; leu2-3,112; ade2-101; his3-11; trp1-1; ura3-1*) transformed with plasmid pAB20 expressing the activated *RAS2-val*^{19} gene (dominant *RAS2* allele) ("WT"); and in a W303-derived strain lacking both the Rce1p and Afc1p CAAX proteases (*Δrce1, Δafc1*) co-transformed with one of the following *URA3* marked, high-copy (2μ) plasmids: pAB1 (vector control), pAB13 (*hRCE1*), pAB15 (H172A), pAB16 (H174A); and pAB17 (H175A) (as described in **Figure 10**). Percent (%) viability after heat shock treatment is shown for each strain (see **Example 9**).

PRIMER SEQUENCES

Oligonucleotides or "primers" were all purchased from Operon Technologies (Alameda, CA) and consisted of the following sequences: hRCE(1R)-RI (SEQ ID NO:7) CGG AAT TCT CAA TAA TAT GGT GAA AAT GGG C; hRCE(2R)-RI (SEQ ID NO:8) CGG AAT TCG GAT GAA GAG GAA AGC AGT G; hRCE(3F)-RI (SEQ ID NO:9) GCG AAT TCT TCG GTG CCT ACA CTG CTT TC; hRCE(4F)-RI (SEQ ID NO:10) GCG AAT TCC CAT TTT CAC CAT ATT ATT GAG C; hRCE(5R)(SEQ ID NO:11) CTA ACA TGG GCA GCA TAC AGG; hRCE(6R)(SEQ ID NO:12) CTG GTG AAG CGT CGC TTG A; hRCE(7F)(SEQ ID NO:13) ATC CAG CCA GGC ACA TCC; hRCE(8F)(SEQ ID NO:14) TTG CCT CAC AGA CAT GCG TTG; hRCE(9F)(SEQ ID NO:15) AAC GCA TGT CTG TGA GGC AAC; hRCE(10R)-RI (SEQ ID NO:16) CGG AAT TCG CTA ACA TGG GCA GCA TAC AGG CC; hRCE(11R)(SEQ ID NO:17) GGT CAG GAG CAC AGG GGA GC; hRCE(12F)(SEQ ID NO:18) GAA GGG CTG GCT GGG GTC; hRCE(13R)(SEQ ID NO:19) GGT CCC TGG GAC GCA ACT AG;

hRCE(14F)(SEQ ID NO:20) CGG GGC GGC GCT GGG CGG GGA TGG GC; hRCE(15R)(SEQ ID NO:21) AGG CTG CCC ACG TAG GAG CAG GCG AGG CTG A, hRCE-ValX (SEQ ID NO:22) GTG CTC TAG AGT GTC AGT GTT CTC CTG CCT CAG CCT C; hRCE3P-RI (SEQ ID NO:23) CCG GGA ATT CAG GAG CAC AGG GGA GCT TGT GAG TCC C; hRCE3P-Xba (SEQ ID NO:24) CCG GTC TAG AGG AGC ACA GGG GAG CTT GTG AGT CCC; 608(1F)-RI (SEQ ID NO:25) CGG AAT TCC CTG CTG TGT TCA CCT GCC CA; 608(2F)-RI (SEQ ID NO:26) CGG AAT TCC TCC TAC ACC GCT GTC TTC GGT G; 608(1R)-RI (SEQ ID NO:27) CGG AAT TCC CAG CGC ATG TCT GTG AGG CAG; 608(2R)-RI (SEQ ID NO:28) CGG AAT TCT GGG CAG GTG AAC ACA GCA GG; hAFC(1R)-RI (SEQ ID NO:29) GCG AAT TCC TGG AAG ATG ATC AAT AGT CCA AT; hAFC(2R)(SEQ ID NO:30) TCT GGG CTT TTC TGT ACT TAA TT; hAFC(3F)(SEQ ID NO:31) GGA ACA GTG AAG AAA TAA AAG CTA AAG; hAFC(4F)(SEQ ID NO:32) TGA GGA GGT ACT CGC TGT ACT AGG; hSTE14(1F)-RI (SEQ ID NO:33) GCG AAT TCG AGA ACA TCT TCT GGC CAG, hSTE14(2R)-RI (SEQ ID NO:34) GCG AAT TCT GGC CAG AAG ATG TTC TCG; hSTE14(3F)(SEQ ID NO:35) CTA CCA GAT AGC CAT CAG AGC; hSTE14(4R)(SEQ ID NO:36) GCT CTG ATG GCT ATC TGG TAG.

Definitions

Unless otherwise defined, all technical and scientific terms used herein have the meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. The practice of the present invention employs, unless otherwise indicated, conventional techniques of chemistry, molecular biology, microbiology, recombinant DNA, genetics and immunology. See, e.g., Maniatis et al., 1982; Sambrook et al., 1989; Ausubel et al., 1992; Glover, 1985; Anand, 1992; Guthrie and Fink, 1991 (which are incorporated herein by reference).

An "isolated" or "substantially pure" nucleic acid, polynucleotide (e.g., an RNA, DNA or a mixed polymer) or polypeptide is one which is substantially

separated from other cellular components that naturally accompany the native sequence or protein in its natural host cell, e.g., ribosomes, polymerases, many other mammalian genome sequences and proteins. The term embraces a nucleic acid sequence or polypeptide that has been removed from its naturally occurring environment, and includes recombinant or cloned DNA isolates and chemically synthesized analogs or analogs that are biologically synthesized by heterologous systems.

The term "nucleic acids" or "polynucleotides" is also extended to refer to nucleic acids that encode a RCE1 or AFC1 polypeptide, polypeptide fragment, homologs and variants of those polypeptides, and protein fusions and deletions of any of the above. These nucleic acids comprise a sequence which is either derived from, or substantially similar to a natural CAAX protease-encoding gene or one having substantial homology with a natural CAAX protease-encoding gene or a portion thereof.

The polynucleotide compositions of this invention may include both sense and antisense strands of RNA, cDNA, genomic DNA, and synthetic forms and mixed polymers of the above. They may be modified chemically or biochemically or may contain non-natural or derivatized nucleotide bases, as will be readily appreciated by those of skill in the art. Such modifications include, for example, labels, methylation, substitution of one or more of the naturally occurring nucleotides with an analog, internucleotide modifications such as uncharged linkages (e.g., methyl phosphonates, phosphotriesters, phosphoramidates, carbamates, etc.), charged linkages (e.g., phosphorothioates, phosphorodithioates, etc.), pendent moieties (e.g., polypeptides), intercalators (e.g., acridine, psoralen, etc.), chelators, alkylators, and modified linkages (e.g., alpha anomeric nucleic acids, etc.) Also included are synthetic molecules that mimic polynucleotides in their ability to bind to a designated

sequence via hydrogen bonding and other chemical interactions. Such molecules are known in the art and include, for example, those in which peptide linkages substitute for phosphate linkages in the backbone of the molecule.

The term "substantial homology or similarity" when referring to a nucleic acid or fragment thereof indicates that, when optimally aligned with appropriate nucleotide insertions or deletions with another nucleic acid (or its complementary strand), there is nucleotide sequence identity in at least about 60% of the nucleotide bases, usually at least about 70%, more usually at least about 80%, preferably at least about 90%, and more preferably at least about 95-98% of the nucleotide bases.

Alternatively, substantial homology or similarity exists when a nucleic acid or fragment thereof hybridizes to another nucleic acid, to a strand of another nucleic acid, or to the complementary strand thereof, under selective hybridization conditions. Typically, selective hybridization will occur when there is at least about 55% homology -- preferably at least about 65%, more preferably at least about 75%, and most preferably at least about 90% -- over a stretch of at least about 14 nucleotides. See, e.g., Kanehisa, 1984. The length of homology comparison, as described, may be over a stretch of at least about nine nucleotides, usually at least about 20 nucleotides, more usually at least about 24 nucleotides, typically at least about 28 nucleotides, more typically at least about 32 nucleotides, and preferably at least about 36 or more nucleotides.

Nucleic acid hybridization will be affected by such conditions as salt concentration, temperature, solvents, the base composition of the hybridizing species, length of the complementary regions, and the number of nucleotide base mismatches between the hybridizing nucleic acids, as will be readily appreciated by those skilled in

the art.

"Stringent hybridization" and "stringent hybridization wash conditions" in the context of nucleic acid hybridization experiments differ depending on different physical parameters such as temperature, the base composition and concentration of the reactants and the salt concentration. Stringent temperature conditions will generally include temperatures in excess of 30°C, typically in excess of 37°C, and preferably in excess of 45°C. Stringent salt conditions will ordinarily be less than 1000 mM, typically less than 500 mM, and preferably less than 200 mM. However, the combination of parameters is much more important than the measure of any single parameter. See, e.g., Wetmur and Davidson, 1968. Probe sequences may also hybridize specifically to duplex DNA under certain conditions to form triplex or other higher order DNA complexes. The preparation of such probes and suitable hybridization conditions are well known in the art.

Generally, stringent hybridization and wash conditions are selected to be about 5°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH. The T_m is the temperature at which 50% of the target sequence hybridizes to a perfectly matched probe. An example of stringent hybridization conditions for hybridization of complementary nucleic acid sequences having more than 100 complementary residues on a filter in a Southern or northern blot is 50% formamide with 1 mg of heparin per 50mL at 42°C for at least ten hours. An example of stringent wash conditions is a 0.2x SSC wash at 65°C for 15 minutes (see Sambrook et al., supra, for SSC buffer). Often the high stringency wash is preceded by a low stringency wash to remove excess probe. An exemplary medium stringency wash for duplex DNA of more than 100 base pairs is 1xSSC at 45°C for 15 minutes. An exemplary

low stringency wash for such a duplex is 4xSSC at 40°C for 15 minutes. In general, signal-to-noise ratio of 2x or higher than that observed for an unrelated probe in the particular hybridization assay indicates detection of a specific hybridization.

As defined herein, nucleic acids which do not hybridize to each other under stringent conditions are still substantially identical to those which do hybridize under stringent conditions if the polypeptides which they encode are substantially identical. This occurs, for example, when a nucleic acid is created synthetically or recombinantly using a high codon degeneracy as permitted by the redundancy of the genetic code.

The term "identical" in the context of nucleic acid or polypeptide sequences refers to the residues in the two sequences which are the same when aligned for maximum correspondence. When percentage of sequence identity is used in reference to proteins or peptides, it is recognized that residue positions which are not identical often differ by conservative amino acid substitutions -- where amino acid residues are substituted for other amino acid residues having side chains with similar chemical properties (e.g., charge or hydrophobicity) which do not substantially change the functional properties of the molecule. Where sequences differ in conservative substitutions the percent sequence identity may be adjusted upwards to correct for the conservative nature of the substitution. Means for making this adjustment are well known to those of skill in the art (see, e.g., Pearson et al., Methods in Molecular Biology, 24, pp. 307-31 (1994)).

The terms "substantial homology" or "substantial identity", when referring to polypeptides, indicate that the polypeptide or protein in question exhibits at least about 30% identity with an entire naturally-occurring protein or a portion thereof, usually at least about 70%

identity, and preferably at least about 95% identity.

The term "substantially similar function" refers to the function of a modified nucleic acid or a modified polypeptide (or protein) with reference to its wild-type nucleic acid or polypeptide counterpart. The modified CAAX protease polypeptide will be substantially homologous to the wild-type CAAX protease polypeptide and will have substantially the same function. The modified polypeptide may have an altered amino acid sequence and/or may contain modified amino acids.

In addition to the similarity of function, a modified CAAX protease polypeptide may have other useful properties, such as an altered (e.g., longer) half-life. The similarity of function (activity) of the modified polypeptide may be substantially the same as the activity of the wild-type CAAX protease polypeptide. Alternatively, the similarity of function (activity) of the modified polypeptide may differ from the activity of the wild-type polypeptide in, for example, substrate specificity and/or enzymatic activity. The modified polypeptide is synthesized using conventional techniques, or is encoded by a modified nucleic acid and produced using conventional techniques. The modified nucleic acid is prepared by conventional techniques. A nucleic acid with a function substantially similar to the wild-type *RCE1* or *AFC1* gene function produces the modified protein as described above.

Homology, for polypeptides, is typically measured using sequence analysis software. See, e.g., the Sequence Analysis Software Package of the Genetics Computer Group, University of Wisconsin Biotechnology Center, 910 University Avenue, Madison, Wisconsin 53705. Protein analysis software matches similar sequences using measure of homology assigned to various substitutions, deletions and other modifications. The length of polypeptide sequences compared for homology will generally

be at least about 16 amino acids, usually at least about 20 residues, more usually at least about 24 residues, typically at least about 28 residues, and preferably more than about 35 residues.

"Conservatively modified variations" of a particular nucleic acid sequence refers to nucleic acids that encode identical or essentially identical amino acid sequences or DNA sequences where no amino acid sequence is encoded. because of the degeneracy of the genetic code, a large number of functionally identical nucleic acids encode any given polypeptide sequence. When a nucleic acid sequence is changed at one or more positions with no corresponding change in the amino acid sequence which it encodes, that mutation is called a "silent mutation". Thus, one species of a conservatively modified variation according to this invention is a silent mutation. Accordingly, every nucleic acid sequence herein which encodes a polypeptide also describes every possible silent mutation or variation.

Furthermore, one of skill in the art will recognize that individual substitutions, deletions, additions and the like, which alter, add or delete a single amino acid or a small percentage of amino acids (less than 5%, more typically less than 1%) in an encoded sequence are "conservatively modified variations" where the alterations result in the substitution of one amino acid with a chemically similar amino acid. Conservative substitution tables providing functionally similar amino acids are well known in the art. The following six groups each contain amino acids that are conservative substitutions for one another:

1) Alanine (A), Serine (S), Threonine (T);
2) Aspartic Acid (D), Glutamic Acid (E);
3) Asparagine (N), Glutamine (Q);
4) Arginine (R), Lysine (K);

5) Isoleucine (I), Leucine (L), Methionine (M), Valine (V), and

6) Phenylalanine (F), Tyrosine (Y), Tryptophan (W).

The terms "RCE1 (or AFC1) protein" or "RCE1 (or AFC1) polypeptide" refer to a protein or polypeptide encoded by the *RCE1* (or *AFC1*) locus, and variants or fragments thereof. The term "polypeptide" refers to a polymer of amino acids and its equivalent and does not refer to a specific length of the product; thus, peptides, oligopeptides and proteins are included within the definition of a polypeptide. This term also does not exclude modifications of the polypeptide, for example, glycosylations, acetylations, phosphorylations, and the like. Included within the definition are, for example, polypeptides containing one or more analogs of an amino acid (including, e.g., unnatural amino acids), polypeptides with substituted linkages as well as other modifications known in the art, both naturally and non-naturally occurring. Ordinarily, such polypeptides will be at least about 50% homologous to the native CAAX protease sequences, preferably in excess of about 90%, and more preferably at least about 95% homologous. Also included are proteins encoded by DNA which hybridize under high or low stringency conditions, to CAAX protease-encoding nucleic acids and closely related polypeptides or proteins retrieved by antisera raised against such protein(s).

A polypeptide "fragment", "portion" or "segment" refers to a stretch of amino acid residues of at least about five to seven contiguous amino acids, often at least about seven to nine contiguous amino acids, typically at least about nine to 13 contiguous amino acids and, most preferably, at least about 20 to 30 or more contiguous amino acids.

The term "protein modifications" refers to RCE1 (or AFC1) polypeptides or fragments thereof that are substantially homologous in primary structural sequence but which include, e.g., in vivo or in vitro chemical and biochemical modifications or which incorporate unusual amino acids. Such modifications include, for example, acetylation, carboxylation, phosphorylation, glycosylation, ubiquitination, labeling, e.g., with radionuclides, and various enzymatic modifications, as will be readily appreciated by those well skilled in the art. A variety of methods for labeling polypeptides and of substituents or labels useful for such purposes are well known in the art, and include radioactive isotopes such as ^{32}P, ligands which bind to labeled antiligands (e.g., antibodies), fluorophores, chemiluminescent agents, enzymes, and antiligands which can serve as specific binding pair members for a labeled ligand. The choice of label depends on the sensitivity required, ease of conjugation with the primer, stability requirements, and available instrumentation. Methods for labeling polypeptides are well known in the art. See Sambrook et al., 1989 or Ausubel et al., 1992.

The term "fusion protein" refers to fusion polypeptides comprising RCE1 or AFC1 polypeptides and fragments. Homologous polypeptide fusions may be between two or more polypeptide sequences or between sequences derived from RCE1p or AFC1p and a related protein. Heterologous fusions may be constructed which would exhibit a combination of properties or activities of the derivative polypeptides. For example, active site or other domains may be "swapped" between CAAX protease polypeptides and other polypeptides or fragments. Such homologous or heterologous fusion polypeptides may display altered biological characteristics (e.g., altered enzymatic activity, substrate specificity and/or altered cellular localization).

The terms "isolated", "substantially pure", "purified", "purified and isolated" and "substantially homogeneous" are used interchangeably to describe a protein or polypeptide that has been separated from components that accompany it in its natural state. A monomeric protein is substantially pure when at least about 60 to 75% of a sample exhibits a single polypeptide sequence. A substantially pure protein will typically comprise about 60 to 90% W/W of a protein sample, more usually about 95%, and preferably will be over 99% pure. Protein purity or homogeneity may be indicated by a number of means well known in the art, such as polyacrylamide gel electrophoresis of a protein sample, followed by visualizing a single polypeptide band upon staining the gel with a stain well known in the art. For certain purposes, higher resolution may be provided by using HPLC or other means well known in the art for purification.

A RCE1 or AFC1 protein is substantially free of naturally associated components when it is separated from the native contaminants that accompany it in its natural state. Thus, a polypeptide that is chemically synthesized or synthesized in a cellular system different from the cell from which it naturally originates will be substantially free from its naturally associated components. A protein may also be rendered substantially free of naturally associated components by isolation, using protein purification techniques well known in the art.

Besides substantially full-length polypeptides, the present invention provides for biologically active fragments of the polypeptides. Significant biological activities include CAAX substrate binding, protease activity, immunological activity or other (presently unidentified) biological activities characteristic of the CAAX protease polypeptides of this invention. Immunological activities include both immunogenic function

in a target immune system, as well as sharing of immunological epitopes for binding, serving as either a competitor or substitute antigen for an epitope of the RCE1 or AFC1 protein.

Expression of recombinant DNA molecules according to this invention may involve post-translational modification of a resultant polypeptide by the host cell. For example, in mammalian cells expression might include, among other things, glycosylation, lipidation or phosphorylation of a polypeptide, or cleavage of a signal sequence to produce a "mature" protein. Accordingly, the polypeptide expression products of this invention encompass full-length polypeptides and modifications or derivatives thereof, such as glycosylated versions of such polypeptides, mature proteins, polypeptides retaining a signal peptide, truncated polypeptides having comparable biological activity, and the like.

As used herein, "epitope" refers to an antigenic determinant of a polypeptide, i.e., a region of a polypeptide that provokes an immunological response in a host. This region needs not comprise consecutive amino acids. The term epitope is also known in the art as "antigenic determinant". An epitope may comprise as few as three amino acids in a spatial conformation which is unique to the immune system of the host. Generally, an epitope consists of at least five such amino acids, and more usually consists of at least 8-10 such amino acids. Methods for determining the spatial conformation of such amino acids are known in the art.

The term "antibody" refers to a polypeptide encoded by an immunoglobulin gene, genes, or fragments thereof. The immunoglobulin genes include the kappa, lambda, alpha, gamma, delta, epsilon and mu constant regions, as well as a myriad of immunoglobulin variable regions. Light chains are classified as either kappa or lambda. Heavy chains are classified as gamma, mu, alpha,

delta, or epsilon, which in turn define the immunoglobulin classes IgG, IgM, IgA, IgD and IgE, respectively.

Antibodies exist for example, as intact immunoglobulins or as a number of well-characterized fragments produced by digestion with various peptidases. For example, trypsin digests an antibody below the disulfide linkages in the hinge region to produce $F(ab)'_2$, a dimer of Fab which itself is a light chain joined to a V_H-C_H1 by a disulfide bond. The $F(ab)'_2$ may be reduced under mild conditions to break the disulfide linkage in the hinge region thereby converting the $F(ab)'_2$ dimer to a Fab' monomer. The Fab' monomer is essentially an Fab with part of the hinge region. See Fundamental Immunology, Third Edition, W.E. Paul, ed. Raven Press, N.Y. (1993) (incorporated herein by reference) for a detailed description of antibodies and antibody fragments. One of skill in the art recognizes that such Fab' fragments may be synthesized de novo either chemically or using recombinant DNA technology. Thus, as used herein, the term antibody includes antibody fragments produced by the modification of whole antibodies or those synthesized de novo. The antibodies of the present invention are optionally derived from libraries of recombinant antibodies in phage or similar vectors (see, e.g., Huse et al., Science, 246, pp. 1275-81 (1989); Ward et al., Nature, 341, pp. 544-46 (1989); Vaughan et al., Nature Biotech., 14, pp. 309-14 (1996) which are incorporated herein by reference).

DETAILED DESCRIPTION OF THE INVENTION

Identifying Human ESTs For Use as Probes

We performed extensive database searches for DNA sequences capable of encoding protein sequence with significant sequence identity with the protein products of

the yeast *RCE1*, *AFC1*, and *STE14* genes using the search engine XREF (Bassett et al., Trends in Genetics, 11, pp. 372-373 (1995). Three expressed sequence tags (ESTs) were found with extensive amino acid sequence identity to the corresponding yeast genes (**Figures 1A-C**). The human EST NCBI#604642 (Genbank Accession # W96411) showed considerable sequence identity with the yeast *RCE1* gene product, Rce1p, over approximately 90 amino acids (**Figure 1A**). In addition, the EST possessed 3 conserved histidine residues in a tetrapeptide sequence we believe is critical for catalysis (HxHH at positions 194-197 of Rce1p; see **Figure 1A**, fragment 188-202).

The human EST NCBI#482520 (Genbank Accession # N76181) showed 45 percent sequence identity with the product of the *AFC1* gene (Afc1p) over approximately 80 amino acids (**Figure 1B**). The murine EST NCBI#636296 (Genbank Accession # AA022288) sequence showed 36 percent identity with the product of the *STE14* gene (Ste14p) over 67 amino acids.

Isolating human *hRCE* cDNA Homologs

To isolate the human cDNA counterpart to the yeast *RCE1* gene, we screened a λgt11 cDNA library made from a 9-week old human fetus. The library (ATCC #77433) was purchased from the American Type Culture Collection (see **Example 1**). Briefly, phage were plated onto an E. coli host strain, lifted onto nylon membranes and hybridized under stringent conditions to radio-labeled DNA sequences which were isolated from the human EST (ID#358628)(see above). Two phage plaques, 13R and 27R, showed a strong hybridization signal to the human EST (#358628) probe and were purified to homogeneity by two successive rounds of plating and hybridization (**Example 1**).

Sequencing the human *hRCE* cDNA

Phage stocks of the λgt11 clones 13R and 27R were prepared and titered, and DNA was prepared from these two clones (**Example 1**). The λgt11 clone 27R contains a human cDNA insert of approximately 1.4 kB. An E. coli strain (LE392) infected with a purified phage stock of λgt11 clone 27R was deposited as strain designation "27R-λgt11" on May 7, 1997 with the American Type Culture Collection (ATCC) (Rockville, MD) according to the provisions of the Budapest Treaty, and was assigned the ATCC accession number designated: ATCC #209020. All restrictions on the availability to the public of the above ATCC deposit will be irrevocably removed upon the granting of a patent on this application.

The 5' and 3' ends of the λ clones were amplified by PCR as described in **Example 1**. Additional sequence information for the human *hRCE* cDNA was obtained by 5′ and 3′ RACE methods (Siebert et al., Nucleic Acids Research, 23, pp. 1087-1088 (1995)) (see **Example 1**).

The consensus sequence derived from the two human λgt11 cDNAs, 13R and 27R, together with the human RACE PCR products revealed the presence of an open reading frame extending from the 5′ end of the sequence to nucleotide 996 (SEQ ID NO:1) (**Figure 2A** and see below).

5′ RACE reactions were also performed on mouse cDNA and a number of products were isolated and cloned into pCR2.1 (obtained from Invitrogen, Inc.). Two genomic *mRCE* clones, 20 (a 10.5 kbp BamHI fragment in pBluescript SK) and 75 (an 8.5 kbp EcoRI fragment in pbluescript SK) were sequenced using the primer RCE-15R. These mouse clone sequences were assembled into one continuous DNA sequence ("a contig") with the aid of a number of murine ESTs (Genbank ID Nos: AA072190, AA154658, AA168614, W14344, AA154864, AA218396). The continuous cDNA and genomic sequences are shown in **Figures 2B and 2C**. The

mouse *RCE* (*mRCE*) cDNA consensus also revealed the presence of a continuous ORF that extended from the 5′ terminus to nucleotide 1004 (SEQ ID NO:3; SEQ ID NO:5)(see **Figure 2B**). The first ATG codon encountered within the mouse ORF is underlined in **Figure 2C**. The predicted protein sequence of mRCE (SEQ ID NO:4; SEQ ID NO:6) is shown in **Figure 2B**.

The first ATG codon encountered within the human *hRCE* ORF (SEQ ID NO:1) is well into the coding sequence and is located beyond a region of sequence identity with the yeast Rce1p sequence (see **Figure 3**) suggesting that translation begins at an upstream site. The 5′ terminal "G" residue of the *hRCE* cDNA sequence (SEQ ID NO:1) (see **Figure 2A**), which was fused to linker sequences during the construction of the cDNA clone, may be preceded by an "AT" dinucleotide, forming an in-frame ATG codon for methionine. As described above, the corresponding G residue in the mouse *mRCE* sequence forms an ATG codon which was revealed upon obtaining additional genomic sequence information. If that ATG codon in *mRCE* is a functional translational start site, it is likely that the 5' G residue in the human *hRCE* sequence shown in **Figure 2A** is also preceded by an AT dinucleotide and the *hRCE* ORF begins with a methionine residue.

Alternatively, translational initiation may begin at a non-ATG codon. There are two in-frame GTG codons at positions 116-118 and 122-124 (SEQ ID NO:1) which are known to function as translational start sites on eukaryotic mRNAs. Initiation at either of these codons would encode a protein of very similar size and composition as the yeast counterpart.

Human *hRCE* RNA Analysis

To resolve whether the DNA sequence we determined from the λgt11 cDNA clones and numerous 5′ RACE products reflected the bona fide 5′ terminus of the *hRCE*

mRNA, we performed Northern blot analyses and primer extension experiments with human mRNA (**Example 2**).

The Northern blots revealed the presence of a single abundant mRNA species approximately 1400 base-pairs in length. The expression pattern indicated that the *hRCE* gene is expressed in all tissues examined. The highest expression was found in placenta and pancreatic mRNA samples. The lowest expression was found in the brain. The size of the *hRCE* mRNA was consistent with transcription beginning at or within the sequence shown in **Figure 2A.**

To map more precisely the actual 5' ends of the *hRCE* mRNA, we performed primer extension experiments (**Example 2**). The reaction products appeared to be the same when the reaction were performed under different conditions, and were the same from the two tissue types (**Figure 5**). The transcriptional start sites appear to range from -10 to -26 relative to the valine codon GTG at position 116-118 (SEQ ID NO:1) (**Figure 2A**).

Expression of the human *hRCE* cDNA

The most simple interpretation of the above analyses is that the start-site of translation of both the *hRCE* and *mRCE* genes is at the GTG codon located at position 116. To test whether the ORF beginning at position 116 encodes a functional counterpart to the yeast gene, we expressed this portion of the human ORF (SEQ ID NO:1 nucleotide 116 to 997, **Figure 2A**) in a yeast strain deleted for both CAAX protease genes, y*AFC1* and y*RCE1* (Boyartchuk et al., supra) (**Example 3**). The human *hRCE* ORF was cloned into one of two expression vectors (pACA1' or pACA5') under the control of the yeast *MEV1* promoter and an ATG translational initiation codon, or the yeast *ERG12* promoter and the first four codons of the *ERG12*

coding sequence, respectively. The resulting plasmids, pACA7001 (also referred to as pAB14) and pACA7002 (also referred to as pAB13), represent translational fusions between the *MEV1* ORF and the *ERG12* ORF, respectively. Both expression vectors contained the yeast PGK1 transcriptional termination sequence.

E. coli strains transformed with either pAB13 or pAB14 were deposited as strain designations " " and " ", respectively, on June , 1998 with the American Type Culture Collection (ATCC) (Rockville, MD) according to the provisions of the Budapest Treaty, and were assigned the ATCC accession number designated: ATCC # and ATCC # . All restrictions on the availability to the public of the above ATCC deposits will be irrevocably removed upon the granting of a patent on this application.

The yeast a-mating factor is a CAAX-containing protein and must undergo C-terminal proteolysis in order to secrete biologically active pheromone. We tested whether the human *hRCE* expression constructs, pACA7001 and pACA7002, were functional in yeast by their ability to restore production of a-factor in strain JRY5317 deleted for the two endogenous CAAX proteases. Plasmids pACA7001 and pACA7002 were each transformed into the yeast strain JRY5317, an a-mating type strain in which both CAAX protease genes have been deleted (Boyartchuk et al., supra) (**Example 3**). We performed a halo assay by replica plating patches of the yeast transformants onto a lawn of α *sst2* tester lawn. Cells harboring an *sst2* mutation are rendered supersensitive to mating pheromone and undergo a G1 arrest when exposed to pheromone.

Cells harboring either one of the *hRCE* expression plasmids, but not the vector control (vector with no *hRCE* insert), produced significant amounts of biologically active a-factor (**Figure 6**). This experiment

demonstrates that the human *hRCE* cDNA (**Figure 2A**) is the functional human counterpart to the yeast *RCE1* gene and has a conserved substrate specificity extending to yeast a-factor. This experiment also demonstrates that the ORF that begins at the GTG codon (SEQ ID NO:1, nucleotides 116-118) and ends at the TGA stop codon (SEQ ID NO:1, nucleotides 995-997) encodes a functional human CAAX protease.

Mammalian homologs of the yeast *AFC1* gene

Comparison of the yeast *AFC1* gene with the NCBI EST database revealed the presence of several ESTs with significant sequence identity. The Genbank Accession numbers for the ESTs are: AA359058, H02497, AA210930, AA295448, Z43273, N76181, T35312, F11310, R54272, AA01886, AA134125, Z45538, H15202, and N30505. The sequences were aligned and found to contain a continuous ORF from the 5' end to nucleotide 721 (SEQ ID NO:37) (see **Figure 7**). The predicted amino acid sequence (SEQ ID NO:38) shares extensive sequence identity with the yeast Afc1p protein, including the proposed active site sequence HExxH (Boyartchuck et al., supra) which is characteristic of zinc metalloproteases (**Figure 8**). Thus, this sequence likely encodes the functional counterpart in humans to the yeast CAAX protease, Afc1p (**Example 4**).

Protein Sequences Related to Human hRce1p

To identify other known protein sequences related to those comprising human hRCE1p, we performed a BLASTP search of GenBank with a fragment of the human Rce1 protein sequence including the histidine-rich sequence "HxHH" (SEQ ID NO:2, amino acids 172-175). The search revealed sequence similarity with both yeast CAAX proteases, yRce1p and yAfc1p (**Figure 9**). The amino acid sequence of yeast Afc1p has a zinc metalloprotease

consensus sequence (VxxHExxH) (SEQ ID NO: 50) located at positions 294-301. In the alignment shown in **Figure 9**, the last histidine of the zinc metalloprotease consensus sequence of yAfc1p, (H301; histidine at position 301) aligns with the histidine at position 172 (H172) of hRce1p. The sequence similarity of the histidine-rich region of hRce1p to the region of yAfc1p comprising the presumptive zinc-binding site which functions in catalysis suggests that these regions serve similar biochemical functions. These similar histidine-rich regions may be involved in substrate binding and/or catalysis. Given the presence of histidine-rich sequences that are functionally important in both proteins and the diverse utilization of histidine residues by protease enzymes in binding zinc, a role in metal binding and hence, catalysis, is reasonable.

Afc1p in all respects appears to be a zinc-dependent metalloprotease based both upon amino acid sequence analysis as well as enzymatic sensitivity to the zinc chelator, *o*-phenanthroline (Boyartchuk et al., supra). Rce1 has no similar sensitivity to *o*-phenanthroline nor to the chelator EDTA. Reports in the literature support the notion that the mammalian CAAX protease is either a serine or cysteine protease based upon its sensitivity to a farnesylated peptide aldehyde (Ma, Gilbert and Rando, Biochemistry, 32, pp. 2386-2393 (1993)) and to the thiol modifying reagents PCMB and NEM (Hrycyna and Clarke, J. Biol. Chem., 267, pp. 10457-464 (1992); Ma, Chaudhuri and Rando, Biochemistry, 31, pp. 11772-777 (1992); Akopyan et al., BBRC,198 , pp. 787-794 (1994)).

Given the modest sensitivity of the mammalian CAAX endoprotease to thiol-modifying reagents and the similarity of hRce1p to Afc1p (**Figure 9**), it is tempting to speculate that Rce1p is related to the insulin degrading enzyme, IDE1.

IDE1 is a zinc-dependent thiol protease that is a member of a family of such proteases that includes the E. coli protease III (Affholter, J.A., Fried, V.A., Roth, R.A., Science, 242, pp. 1415-1418 (1988)), and the yeast protease Axl1p (Adames, N., Blundell, K., Ashby, M.N., Boone, C., Science, 270, pp. 464-467 (1995)). However, the sequence of the yeast or human Rce1 protein do not show any significant sequence similarity to any members of this family.

Perhaps the greatest clue as to the functional properties of Rce1p comes from the observations that Rce1p is inhibited by Zn^{2+} (Ki~100mM) (Ashby, M.N., King, D.S., Rine, J., Proc. Natl. Acad. Sci. U.S.A., 89, pp. 4613-4617 (1992)), that the conserved histidine residues of hRcep1 (H172, H174, H175) are functionally important (see below), and that the same histidine-rich region shows similarity to Afc1p around its putative zinc-binding site described herein. These latter observations strongly suggest a role of metal-binding in the function of hRce1p. Zinc is the most likely metal ion to associate with hRce1p, although another divalent cation may be involved, such as cobalt, calcium or magnesium.

The notion that hRce1p requires zinc-binding, or another divalent cation, will provide valuable information in the design of potent inhibitors. For instance, hydroxamic acid derivatives have found significant utility for the inhibition of zinc proteases such as the matrix metalloproteinase (Levy, D.E. et al., J. Med. Chem., 41, pp. 199-223 (1998)) and angiotensin-converting enzyme (Parvathy, S. et al., Biochem. J., 327, pp. 37-43 (1997)). Hydroxamic acid confers it inhibitory effect on zinc metalloproteases through its metal chelating properties. Possible hydroxamic acid hRce1p inhibitors would thus include a prenylated cysteine amino acid linked to a hydroxamate group at it carboxy terminus. The amino group of the inhibitor would be preferably linked to a small

aliphatic group such as methyl, or more preferably to acetyl, or linked to a small number of amino acids, or left unmodified. The prenyl group would be linked to the cysteine residue through the sulfur group of the cysteine and would be composed of a *trans,trans,trans*-geranylgeranyl lipid, or more preferably, a *trans,trans*-farnesyl lipid.

Mutagenesis of the Histidine-Rich Region of *hRCE1*

To determine whether the histidine-rich sequence "HxHH" of the human Rce1 protein (SEQ ID NO:2, amino acids 172-175) are required for biological activity, we changed each of these histidine residues individually to an alanine residue (H172A, H174A and H175A) by site-directed mutagenesis (**Example 7**). Each mutant *hRCE1* gene was inserted into a high-copy expression vector and tested in parallel with a wild-type *hRCE1* expression construct for function in yeast by their ability to restore production of a-factor as determined by halo assays in strain JRY5317 (deleted for the two endogenous CAAX proteases) (see **Figure 6** and accompanying text, above, and **Example 3**).

As shown in **Figure 10**, cells harboring the wild-type *hRCE* expression plasmid produced significant amounts of biologically active a-factor producing a visible halo (pAB13; also referred to as pACA7002 in **Figure 6**). In contrast, cells harboring mutant *hRCE* expression plasmids in which the histidines at position 172, 174 or 175 were individually replaced by alanine produced significantly less biologically active a-factor, as evidenced by the reduced (pAB16; H174A and pAB17; H175A) or undetectable (pAB15; H172A) halos (**Figure 10**). This experiment demonstrates that histidine residues H172, H174 and H175, comprising the histidine-rich region of human *hRCE1* (conserved between *RCE1* and *AFC1*) contribute to biological

activity of hRCE1p protein in vivo.

The ability for the above wild-type and mutant *hRCE1* plasmids to express biologically active CAAX protease in vivo was also measured by heat shock assays in yeast cells transformed with a plasmid expressing the mutant *RAS2-val*^{19} gene (dominant *RAS2* allele) (see **Figure 11** and **Example 9**).

As described above, a deletion in the *RCE1* gene suppresses the temperature-sensitive phenotype of cells expressing activated Ras2p-val^{19} (i.e., Rce1p activity is required for the functional expression of Ras2p). Heat-shock assays in cells expressing activated Ras2p-val^{19} may thus be used to assess Rce1p biological activity in vivo (see **Example 9**). As shown in **Figure 11**, cells that express activated Ras2p-val^{19} in the presence of a functional *RCE1* gene (WT; wild-type *yRCE1* and pAB13; *hRCE1*) are temperature-sensitive after heat shock and have reduced assay or with a wild-type Percent (%) viability after heat shock treatment is shown for each strain .

Vectors, Cloning and Nucleic Acids

The present invention provides nucleic acids and recombinant DNA vectors which comprise mammalian *RCE* and *AFC* DNA sequences. More specifically, vectors comprising all or portions of the cDNA sequences of the human and mouse *RCE* and *AFC* functional homologs described above are provided. The vectors of this invention also include those comprising DNA sequences which hybridize under stringent conditions to the human and mouse *RCE* and *AFC* cDNA sequences, and conservatively modified variations thereof.

The nucleic acids of this invention include single stranded and double-stranded genomic DNA, cDNA, RNA, or hybrids thereof and may be isolated from biological sources or synthesized chemically or by

recombinant DNA methodology. The nucleic acids, recombinant DNA molecules and vectors of this invention may be present in transformed or transfected cells, cell lysates, or in partially purified or substantially pure forms.

As is well known in the art, DNA sequences may be expressed by operatively linking them to an expression control sequence in an appropriate expression vector and employing that expression vector to transform an appropriate unicellular host. Such operative linking of a DNA sequence of this invention to an expression control sequence, of course, includes, if not already part of the DNA sequence, the provision of a translation initiation codon, ATG, in the correct reading frame upstream of the DNA sequence.

A wide variety of host/expression vector combinations may be employed in expressing the DNA sequences of this invention. Useful expression vectors, for example, may consist of segments of chromosomal, non-chromosomal and synthetic DNA sequences.

DNA vector design for transfection into mammalian cells should include appropriate sequences to promote expression of the gene of interest, including: appropriate transcription initiation, termination and enhancer sequences; efficient RNA processing signals such as splicing and polyadenylation signals; sequences that stabilize cytoplasmic mRNA; sequences that enhance translation efficiency (i.e., Kozak consensus sequence); sequences that enhance protein stability; and when desired, sequences that enhance protein secretion. A great number of expression control sequences -- constitutive, inducible and/or tissue-specific -- are known in the art and may be utilized.

Preferred DNA vectors also include a marker gene and means for amplifying the copy number of the gene of interest. DNA vectors may also comprise stabilizing

sequences (e.g., ori- or ARS-like sequences and telomere-like sequences), or may alternatively be designed to favor directed or non-directed integration into the host cell genome.

Substantial progress in the development of mammalian cell expression systems has been made in the last decade and many aspects of the system are well characterized. A detailed review of the production of foreign proteins in mammalian cells, including useful cells, protein expression-promoting sequences, marker genes, and gene amplification methods, is disclosed in M. M. Bendig, Genetic Engineering, 7, pp. 91-127 (1988).

Particular details of the transfection, expression and purification of recombinant proteins are well documented and are understood by those of skill in the art. Further details on the various technical aspects of each of the steps used in recombinant production of foreign genes in mammalian cell expression systems can be found in a number of texts and laboratory manuals in the art. See, e.g., F. M. Ausubel et al., ed., Current Protocols in Molecular Biology, John Wiley & Sons, New York (1989).

Useful expression vectors for bacterial hosts include bacterial plasmids, such as those from *E. coli*, including pBluescript, pGEX-2T, pUC vectors, col E1, pCR1, pBR322, pMB9 and their derivatives, wider host range plasmids, such as RP4, phage DNAs, e.g., the numerous derivatives of phage lambda, e.g., NM989, λGT10 and λGT11, and other phages, e.g., M13 and Filamenteous single stranded phage DNA. Useful expression vectors for yeast cells include the 2μ plasmid and derivatives thereof. Useful vectors for insect cells include pVL 941.

In addition, any of a wide variety of expression control sequences -- sequences that control the expression of a DNA sequence when operatively linked to it -- may be used in these vectors to express the DNA sequences of this

invention. Such useful expression control sequences include the expression control sequences associated with structural genes of the foregoing expression vectors. Examples of useful expression control sequences include, for example, the early and late promoters of SV40 or adenovirus, the lac system, the trp system, the TAC or TRC system, the T3 and T7 promoters, the major operator and promoter regions of phage lambda, the control regions of fd coat protein, the promoter for 3-phosphoglycerate kinase or other glycolytic enzymes, the promoters of acid phosphatase, e.g., Pho5, the promoters of the yeast α-mating system, the GAL1 or GAL10 promoters, and other constitutive and inducible promoter sequences known to control the expression of genes of prokaryotic or eukaryotic cells or their viruses, and various combinations thereof. See, e.g., The Molecular Biology of the Yeast Saccharomyces (eds. Strathern, Jones and Broach) Cold Spring Harbor Lab., Cold Spring Harbor, N.Y. for details on yeast molecular biology in general and on yeast expression systems (pp. 181-209) (incorporated herein by reference)).

For eukaryotic cells, expression control sequences typically include a promoter, an enhancer derived from immunoglobulin genes, SV40, cytomegalovirus, etc., and a polyadenylation sequence which may include splice donor and acceptor sites.

In a preferred embodiment, DNA sequences of this invention are cloned in the expression vector lambda ZAP II (Stratagene, La Jolla, CA), in which expression from the lac promoter may be induced by IPTG.

In another preferred embodiment, DNA sequences of this invention are inserted in frame into an expression vector that allows high level expression of the corresponding polypeptide as a fusion protein. Examples of fusion proteins envisioned by this invention include but are not limited to fusions to reporter genes, such as

the green fluorescent protein, lacZ and luciferase, and fusions to epitope tags such as myc, hemagglutinin (HA) or other related short immunogenic sequences. The DNA sequences of this invention may be linked by the acid labile cleavage site (Asp-Pro) to a leader sequence (such as the "MLE leader") suitable for promoting expression in *E. coli*. Full-length and truncated forms of the DNA sequences of this invention may be prepared as fusion constructs.

A wide variety of unicellular host cells are useful in expressing the DNA sequences of this invention. These hosts may include well known eukaryotic and prokaryotic hosts, such as strains of *E. coli*, *Pseudomonas*, *Bacillus*, *Streptomyces*, fungi, yeast, insect cells such as *Spodoptera frugiperda* (SF9), animal cells such as CHO and mouse cells, African green monkey cells such as COS 1, COS 7, BSC 1, BSC 40, and BMT 10, and human cells, as well as plant cells in tissue culture.

It should of course be understood that not all vectors and expression control sequences will function equally well to express the DNA sequences of this invention. Neither will all hosts function equally well with the same expression system. However, one of skill in the art may make a selection among these vectors, expression control sequences and hosts without undue experimentation and without departing from the scope of this invention. For example, in selecting a vector, the host must be considered because the vector must be replicated in it. The vector's copy number, the ability to control that copy number, the ability to control integration, if any, and the expression of any other proteins encoded by the vector, such as antibiotic or other selection markers, should also be considered.

In selecting an expression control sequence, a variety of factors should also be considered. These

include, for example, the relative strength of the sequence, its controllability, and its compatibility with the DNA sequence of this invention, particularly with regard to potential secondary structures. Unicellular hosts should be selected by consideration of their compatibility with the chosen vector, the toxicity of the product coded for by the DNA sequences of this invention, their secretion characteristics, their ability to fold the polypeptide correctly, their fermentation or culture requirements, and the ease of purification from them of the products coded for by the DNA sequences of this invention.

Within these parameters, one of skill in the art may select various vector/expression control sequence/host combinations that will express the DNA sequences of this invention on fermentation or in other large scale cultures.

Given the strategies described herein, one of skill in the art can construct a variety of vectors and nucleic acid molecules comprising functionally equivalent nucleic acids. DNA cloning and sequencing methods are well known to those of skill in the art and are described in an assortment of laboratory manuals, including Sambrook et al, supra, 1989; and Current Protocols in Molecular Biology, F.M. Ausubel et al., eds., Current Protocols, a joint venture between Greene Publishing Associates, Inc. and John Wiley & Sons, Inc. (1994 Supplement). Product information from manufacturers of biological, chemical and immunological reagents also provide useful information.

Recombinant Mammalian RCE and AFC Polypeptides and Antibodies Directed Against Them.

The recombinant DNA molecules and more particularly, the expression vectors of this invention may be used to express recombinant RCE and AFC polypeptides in

a heterologous host cell. The RCE and AFC polypeptides of this invention may be full-length or less than full-length recombinant polypeptide fragments recombinantly expressed from the DNA sequences according to this invention. Such polypeptides include variants, fragments and muteins having biological activity. The polypeptides of this invention may be soluble, or may be engineered to be membrane- or substrate-bound using techniques (such as cross-linking) well known in the art. The polypeptides of this invention may be fused to other molecules, such as genetic, enzymatic or chemical or immunological markers such as epitope tags. Fusion partners include, inter alia, GST, immunoglobulins, bacterial beta-galactosidase, trpE, protein A, β-lactamase, α amylase, maltose binding protein, alcohol dehydrogenase, polyhistidine (for example, six histidine at the amino and/or carboxyl terminus of the polypeptide), green fluorescent protein (GFP), yeast α mating factor, GAL4 transcription activation or DNA binding domain, and luciferase. See, e.g., Godowski et al., 1988.

Fusion proteins will typically be made by either recombinant nucleic acid methods, as described above, or may be chemically synthesized. Techniques for the synthesis of polypeptides are described, for example, in Merrifield, 1963.

Transformation and other methods of introducing nucleic acids into a host cell (e.g., transfection and viral infection) can be accomplished by a variety of methods which are well known in the art. Animal (and yeast) cells may be transfected with a plasmid, a cosmid or the like, infected with a viral vector of this invention for transient or stable expression of RCE or AFC polypeptides. For example, for transient expression in animal cells, investigators have used spheroplast fusion, DEAE dextran, and electroporation. For stable expression they have used calcium phosphate, spheroplast fusion, and

electroporation.

The polypeptide gene products of the mammalian *RCE* or *AFC* genes of this invention may be used to produce polyclonal or monoclonal antibodies using a variety of techniques well known to those of skill in the art. Alternatively, peptides corresponding to specific regions of the RCE or AFC polypeptides may be synthesized and used to create immunological reagents according to well known methods.

Antibodies directed against the mammalian RCE and AFC polypeptides of this invention are immunoglobulin molecules or portions thereof that are immunologically reactive with at least one novel mammalian RCE or AFC polypeptide of the present invention. It should be understood that the antibodies of this invention include antibodies immunologically reactive with fusion proteins and multimeric proteins comprising a novel mammalian RCE or AFC polypeptide.

Antibodies directed against a novel mammalian RCE or AFC polypeptide may be generated by immunization of a mammalian host. Such antibodies may be polyclonal or monoclonal. Preferably they are monoclonal. Methods to produce polyclonal and monoclonal antibodies are well known to those of skill in the art. For a review of such methods, see *Antibodies, A Laboratory Manual, supra*, and D.E. Yelton, et al., Ann. Rev. of Biochem., 50, pp. 657-80 (1981). Determination of immunoreactivity with a novel mammalian RCE or AFC polypeptide of this invention may be made by any of several methods well known in the art, including by immunoblot assay and ELISA.

Monoclonal antibodies with affinities of 10^{-8} M^{-1} or preferably 10^{-9} to 10^{-10} M^{-1} or stronger are typically made by standard procedures as described, e.g., in Harlow and Lane, 1988 or Goding, 1986. Briefly, appropriate animals are selected and the desired

immunization protocol followed. After the appropriate period of time, the spleens of such animals are excised and individual spleen cells fused, typically, to immortalized myeloma cells under appropriate selection conditions. Thereafter, the cells are clonally separated and the supernatants of each clone tested for their production of an appropriate antibody specific for the desired region of the antigen.

Other suitable techniques involve in vitro exposure of lymphocytes to the antigenic polypeptides, or alternatively, to selection of libraries of antibodies in phage or similar vectors. See Huse et al., 1989. The polypeptides and antibodies of the present invention may be used with or without modification. Frequently, polypeptides and antibodies will be labeled by joining, either covalently or non-covalently, a substance which provides for a detectable signal. A wide variety of labels and conjugation techniques are known and are reported extensively in both the scientific and patent literature. Suitable labels include radionuclides, enzymes, substrates, cofactors, inhibitors, fluorescent agents, chemiluminescent agents, magnetic particles and the like. Patents teaching the use of such labels include U.S. Patents 3,817,837; 3,850,752; 3,939,350; 3,996,345; 4,277,437; 4,275,149 and 4,366,241. Also, recombinant immunoglobulins may be produced (see U.S. Patent 4,816,567).

An antibody of this invention may also be a hybrid molecule formed from immunoglobulin sequences from different species (e.g., mouse and human) or from portions of immunoglobulin light and heavy chain sequences from the same species. It may be a molecule that has multiple binding specificities, such as a bifunctional antibody prepared by any one of a number of techniques known to those of skill in the art including: the production of hybrid hybridomas; disulfide exchange; chemical cross-

linking; addition of peptide linkers between two monoclonal antibodies; the introduction of two sets of immunoglobulin heavy and light chains into a particular cell line; and so forth.

The antibodies of this invention may also be human monoclonal antibodies, for example those produced by immortalized human cells, by SCID-hu mice or other non-human animals capable of producing "human" antibodies, or by the expression of cloned human immunoglobulin genes.

In sum, one of skill in the art, provided with the teachings of this invention, has available a variety of methods which may be used to alter the biological properties of the antibodies of this invention including methods which would increase or decrease the stability or half-life, immunogenicity, toxicity, affinity or yield of a given antibody molecule, or to alter it in any other way that may render it more suitable for a particular application.

Assays to Identify Inhibitors of Mammalian CAAX-Processing Enzymes

This invention is directed to methods for identifying inhibitors of mammalian CAAX processing enzymes by <u>in vitro</u> enzyme inhibition assays using the polypeptide expression products of the mammalian *RCE* and *AFC* sequences. In particular, this invention provides a heterologous host cell or organism comprising a recombinant DNA molecule (vector) capable of expressing a mammalian gene encoding Rce1p or Afc1p. A host cell that overexpresses Rce1p or Afc1p is also provided. One preferred host organism is a yeast such as Saccharomyces transformed with a high copy plasmid encoding a *RCE* or *AFC* gene. The resulting recombinant organism represents an enriched source for the corresponding mammalian enzyme. Inclusion of such an extract in an <u>in vitro</u> CAAX-protease

assay allows the identification of inhibitors that can reduce or eliminate the mammalian CAAX protease activity.

One method for isolating a protein inhibitor (or an activator) of a mammalian CAAX protease is the "two-hybrid system" (see, e.g., C.T. Chien et al., "The two-hybrid system: a method to identify and clone genes for proteins that interact with a protein of interest", Proc. Natl. Acad. Sci. U.S.A., 88(21), pp. 9578-82 (1991)) (incorporated herein by reference) (**Example 6**).

This invention is also directed to methods for identifying inhibitors of mammalian CAAX processing enzymes by screening host cells transformed with the DNA sequences of this invention in in vivo biological activity assays, including but not limited to those described herein.

One preferred in vivo method to identify inhibitors of Rce1p or Afc1p first requires identifying genes whose expression requires the presence of the appropriate protease. A gene expression reporter construct may be created and introduced into cells. Such a cell line or strain will allow indirect monitoring of protease function. Incubation of these cells with select candidate protease inhibitors or a combination of candidate inhibitors such as a chemical library facilitates identification of chemicals that alter protease activity in vivo.

Another preferred in vivo method to identify inhibitors of Rce1p or Afc1p capitalizes on the extreme sensitivity of co-culture experiments to identify differences in the relative growth rates (or fitness) of organisms. Using this approach, two strains of cells which may possess only a subtle difference are mixed and grown for a number of generations. At various points of the growth period, portions of the co-culture are analyzed to determine the relative proportion of each constituent strain. Very small differences in growth rates between

two strains will result in the more fit strain becoming the predominant, if not the exclusive, strain in the co-culture after a number of generations. Co-cultures may be grown and the cells diluted into fresh medium to continually increase the number of cell generations in culture, enabling even very subtle differences in growth rates to be measured.

Such co-culture experiments may be used to identify compounds that are competitive inhibitors of Rce1p or Afc1p (**Example 8**). Competitive inhibition of an enzyme can be overcome by increasing the amount of enzyme. Therefore, a strain that harbors a significant amount of Rce1p (or Afc1p) will be less sensitive to the effects of a competitive inhibitor than will be a strain with a normal amount of Rce1p (or Afc1p).

In one embodiment, a yeast strain is constructed (Strain 1) that harbors a high-copy (2μ) plasmid that encodes the *RCE1* gene. Preferably, the *RCE1* gene is driven by a very strong promoter, such as that of the *ADH1* gene. This *RCE1* overexpressing strain also possesses, preferably (but not necessarily) on the same high-copy plasmid, a marker gene the presence of which (i.e., the gene or gene product) can be easily detected, such as the green fluorescent protein (GFP). A second strain (Strain 2) is required for the co-culture experiment and is preferably Strain 1 without the high-copy plasmid carrying the *RCE1* gene (and optionally comprising a marker gene).

Strain 1 and Strain 2 are grown independently to mid-log phase and then mixed at a selected ratio to create an appropriate co-culture. Preferably, the mixture will contain a large excess of Strain 2 relative to Strain 1 (e.g., 5,000:1; see **Example 8**). Because Strain 1 (which harbors the multi-copy plasmid encoding *RCE1* and GFP) is present in such low abundance, the overall mixture will not exhibit significant fluorescence. During growth of

the co-culture, if a situation were encountered where Strain 1 has a growth advantage over Strain 2, the relative abundance of Strain 1 would increase and thereby cause an easily identified increase in fluorescence of the mixture.

The co-cultures are aliquoted into separate growing chambers (e.g., the wells of 96-well plates). To identify inhibitors of Rcelp, a selected concentration or range of concentrations from selected putative inhibitors (e.g., portions of a chemical library) are added to individual wells of 96-well plates which contain aliquots of the co-culture. Both of these steps lend themselves well to automation and large numbers of chemicals may be tested in parallel.

The co-cultures containing added test chemicals are incubated until the cells have divided numerous times. At various points during the growth period, the relative abundance of Strain 1 in the mixture may be assessed by fluorescence emission from the 96-well plates.

The cells are diluted into fresh medium and allowed to continue to grow. This regimen of growth followed by dilution into fresh medium may be repeated until the cells have experienced multiple generation doublings (e.g., 2-20). Preferably, the cells undergo 10 generation doublings. More preferably, the cells undergo 5 generation doublings. The preferred number of generation doublings of the co-culture may be determined as described in **Example 8**. Preferably, the concentrations and ratios of Strains 1 and 2 in the beginning co-culture are selected to avoid the need to dilute the cultures, which involves more work and time.

In a more preferred embodiment of the above assay, Strain 1 and Strain 2 do not express (e.g., are deleted for) the endogenous *RAS1* gene and possess the *ras2-23* allele in place of *RAS2*. The *ras2-23* allele, for

example, is temperature sensitive and does not adequately function at 37°C to confer viability to a *ras1::HIS3* strain (H. Mitsuzawa et al., Genetics, 123, pp. 739-748 (1989)). At a semi-permissive temperature (34°C), growth of this strain depends upon the function of Rcelp (V.L. Boyartchuk et al., Science, 275, pp. 1796-1800 (1997)). Therefore, the presence of the *ras2-23* allele, in the absence of RAS1, sensitizes Strain 1 and Strain 2 such that their growth rates depend upon the function of Rcelp.

Utility of Mammalian *RCE* and *AFC* Nucleic Acid Sequences, Polypeptides, and Antibodies And Inhibitors Thereof

The sequences of *hRCE* (SEQ ID NO:1), *mRCE* (SEQ ID NO:3 or SEQ ID NO:5) and *hAFC* (SEQ ID NO:37) -- in their entirety or fragments thereof -- may be used in hybridization experiments to: A) determine *RCE* expression levels in various tissues, different developmental or disease states or other conditions; B) isolate and clone related genes, mRNA or cDNAs which hybridize to those sequences; and C) isolate related or identical mRNA sequences for the purpose of in vitro translation. The sequences comprising *hRCE*, *mRCE* or *hAFC* may also be used to construct fusion proteins in order to address the localization of each of these proteins in vivo. Such fusion proteins may additionally contain a reporter gene, such as the green fluorescent protein, lacZ or luciferase.

Alternatively, such sequences may be fused to an epitope tag, such as myc, hemagglutinin (HA) or any other related short sequences. The above DNA sequence manipulations may be accomplished using methods which are well known to those of skill in the art.

RCE and *AFC* sequences may be used to construct vectors that express protein products having biological activity (e.g., CAAX processing) in a heterologous organism, preferably, in a unicellular organism.

Expression products may be purified by conventional means which are well known to those of skill in the art. Mammalian *RCE* and *AFC* DNA sequences or their expression products may be purified and then immobilized to beads or solid support matrices which can then be used to isolate and to identify cellular factors which bind to those DNA or amino acid sequences.

The *hRCE*, *mRCE* or *hAFC* sequences (in their entirety or fragments thereof) may be expressed in a heterologous organism to:

A) examine many biochemical features of the protein such as catalytic mechanism, Kcat, Km, Vmax, and oligomeric status;

B) determine the substrate specificities of the CAAX processing enzyme such as the tolerance of non-aliphatic residues at the A1 and A2 position, and permissive and non-permissive substitutions at the X position, lipid specificities attached to the cysteine and amino acid sequence specificities N-terminal of the cysteine;

C) create an in vitro assay for the purpose of identifying and characterizing inhibitors of the protease;

D) create antibodies (either polyclonal or monoclonal), and hybrid, chimeric and humanized antibodies, for the purpose of localization of protein or to selectively inactivate protease activity;

E) assess the functional importance of Ras CAAX processing in yeast. Yeast which express an activated allele of Ras display a heat shock sensitivity when exposed to an acute treatment at 55°C. The endogenous protease gene products may be inactivated (e.g., by mutation of the *AFC1* and *RCE1* genes) by standard methods, which causes a suppression of the heat-shock sensitivity. Normal, mutant and altered forms of the *hRCE* or *mRCE* cDNAs may be expressed in yeast cells lacking endogenous AFC1 and RCE1 activities (e.g., the deletion strains described

herein) and the yeast cells tested for relative sensitivity to heat-shock.

In addition, *hRCE*, *mRCE* or *hAFC* sequences may be used to create an antisense vector. The antisense vector may be used to synthesize an *RCE* or *AFC* antisense molecule to specifically inactivate gene expression and thereby to inactivate the production of the RCEp or AFCp proteins. Such vectors could be constructed such that it is operably linked and under the control of an inducible, tissue specific, developmental or constitutive promoter. Thus, *RCE* or *AFC* activity could be selectively turned on or off at desired locations that would therapeutically effect a CAAX processing-mediated or affected diseased state or disorder in a patient.

Similarly, the *hRCE*, *mRCE* or *hAFC* sequences of this invention may be used to construct specific ribozymes that are directed to the endogenous *RCE* or *hAFC* mRNAs. Methods for constructing various types of ribozymes are published and may be performed by those of skill in the art using the *RCE* or *hAFC* sequences described herein.

The mouse *mRCE* cDNA sequences of this invention may be used to construct a genetic knockout in a mouse. The genetically altered mouse may in turn be used as a mammalian model for human diseases and disorders involving proteins processed by the *RCE* gene product. Such genetically altered mice will thus be useful in determining the functional importance of mRCE in various mammalian disease states and disorders, and to assess the therapeutic value of inhibitors of mRCE or other therapeutic agents on the diseased state in mouse. An example of one such diseased state is the tumorigenesis induced in mice by expressing activated forms of Ras protein ($H\text{-}ras^{val12}$ or $K\text{-}ras^{val12}$).

The antibodies of this invention have a variety of uses. For example, they are useful as reagents to

screen for expression of mammalian RCE or AFC polypeptides, either in expression libraries constructed from mammalian DNA sources or from other samples in which the proteins may be present. Moreover, by virtue of their specific binding affinities, the antibodies of this invention are also useful to purify or remove polypeptides from a given sample, and to block or bind to specific epitopes on the polypeptides.

Such immunological reagents may be used for a variety of applications, some examples of which are below.

1. Determine protein levels of Rce1p and Afc1p under different conditions in a particular tissue. The assay may be performed by an immobilization-based protocol such as an immunoblot where the proteins are resolved by SDS-PAGE and transferred to a solid support or by an ELISA assay. Such an approach would reveal steady state levels of the proteins. The synthesis rate and protein stability of the respective proteins could be determined by a pulse-chase experiment comprising an immunoprecipitation step using an antibody of this invention.

2. Determine the apparent molecular weights of Rce1p and Afc1p. An immunoblot of protein extracts that have been resolved by SDS-PAGE may be used to identify the size of the proteins when they are expressed in a particular cell type or tissue source or when they are expressed recombinantly in, for example, a unicellular host. Deviations from the molecular weight predicted from the amino acid sequence may
reveal polypeptide products from alternatively spliced mRNAs or post-translational processing events such as proteolysis, glycosylation, lipidation or phosphorylation.

3. Subcellular localization of RCE and AFC polypeptides. Tissue or cell extracts may be fractionated

into cellular components, the proteins resolved by electrophoresis and detected using the immunological reagents of this invention. Association of the protein with a specific cellular organelle may reveal the sub-cellular localization of the protease. Alternatively, antibodies directed against Rce1p or Afc1p could be labeled with a fluorescent tag such as fluorescein and incubated with permeabilized cells in culture. Fluorescent microscopy could then be employed to identify the localization of the protein in vivo.

4. Determining the functional requirement of Ras C-terminal proteolysis. Micro-injection of activated forms of Ras into NIH-3T3 cells or Xenopus oocytes results in a rapid and quantifiable phenotype for assaying Ras function (Schafer et al., Science, 245, pp. 379-385 (1989)). Co-injecting an Rce1p inactivating antibody with an unprocessed form of activated Ras would allow the determination of the functional requirement of Ras proteolytic processing for Ras biological function.

5. Identifying the topology of Rce1p or Afc1p within a lipid bilayer. Monoclonal antibodies directed against discrete regions of either polypeptide could be incubated with permeabilized cells or membrane fractions to map regions of the respective proteins that are cytosol-facing or cytosol-excluded. Cytosol-facing domains would be accessible for interaction with epitope-specific antibodies and could be identified by fluorescent microscopy or immunoblotting techniques.

6. Confirming protein-protein interactions. Candidate Rce1p- or Afc1p-interacting proteins identified genetically or by an assay such as the "two-hybrid screen" could be tested for their ability to interact in vitro. Cells ectopically expressing the candidate protein and the

respective protease could be cross-linked and immunoprecipitated with protein- specific antisera. Once precipitated, the complexes could be resolved by SDS-PAGE and immunoblotted with antisera directed against the second member of the putative complexes.

All of the above utilities apply to the corresponding use of the *AFC* sequences, polypeptides and inhibitors and antibodies of those polypeptides according to this invention whether or not specifically referred to above. They also apply to the corresponding use of related sequences which may be identified, for example, by hybridization to the sequences of this invention, and their corresponding expression products.

This invention is also directed to methods for identifying inhibitors of mammalian CAAX processing enzymes by screening host cells transformed with the DNA sequences of this invention in <u>in vivo</u> biological activity assays such as those described herein, or by <u>in vitro</u> enzyme inhibition assays using the polypeptide expression products of the mammalian *RCE* and *AFC* sequences. Inhibitors of the mammalian CAAX processing enzymes will be useful as therapeutics for treating certain CAAX protein-mediated diseased states or disorders in a patient, for example, for treating cancers or tumors correlated to Ras activation. In the therapeutic method of this invention, a small molecule inhibitor of a mammalian Rce1p or Afc1p protein is administered to a patient suffering from a CAAX protein-mediated disease or disorder.

The inhibitors of mammalian CAAX processing enzymes of this invention may be formulated into pharmaceutical compositions and administered <u>in vivo</u> at an effective dose to treat the particular clinical condition addressed. Determination of a preferred pharmaceutical

formulation and a therapeutically efficient dose regiment for a given application is within the skill of the art taking into consideration, for example, the condition and weight of the patient, the extent of desired treatment and the tolerance of the patient for the treatment.

Administration of the inhibitors of mammalian CAAX processing enzymes of this invention, including isolated and purified forms, their salts or pharmaceutically acceptable derivatives thereof, may be accomplished using any of the conventionally accepted modes of administration of agents which are used to treat, for example, tumors and cancers.

The pharmaceutical compositions of this invention may be in a variety of forms, which may be selected according to the preferred modes of administration. These include, for example, solid, semi-solid and liquid dosage forms such as tablets, pills, powders, liquid solutions or suspensions, suppositories, and injectable and infusible solutions. The preferred form depends on the intended mode of administration and therapeutic application. Modes of administration may include oral, parenteral, subcutaneous, intravenous, intralesional or topical administration.

The inhibitors of this invention may, for example, be placed into sterile, isotonic formulations with or without cofactors which stimulate uptake or stability. The formulation is preferably liquid, or may be lyophilized powder. For example, the inhibitors may be diluted with a formulation buffer comprising 5.0 mg/ml citric acid monohydrate, 2.7 mg/ml trisodium citrate, 41 mg/ml mannitol, 1 mg/ml glycine and 1 mg/ml polysorbate 20. This solution can be lyophilized, stored under refrigeration and reconstituted prior to administration with sterile Water-For-Injection (USP).

The compositions also will preferably include conventional pharmaceutically acceptable carriers well

known in the art (see for example Remington's Pharmaceutical Sciences, 16th Edition, 1980, Mac Publishing Company). Such pharmaceutically acceptable carriers may include other medicinal agents, carriers, genetic carriers, adjuvants, excipients, etc., such as human serum albumin or plasma preparations. The compositions are preferably in the form of a unit dose and will usually be administered one or more times a day.

The pharmaceutical compositions of this invention may also be administered using microspheres, liposomes, other microparticulate delivery systems or sustained release formulations placed in, near, or otherwise in communication with affected tissues or the bloodstream. Suitable examples of sustained release carriers include semipermeable polymer matrices in the form of shaped articles such as suppositories or microcapsules. Implantable or microcapsular sustained release matrices include polylactides (U.S. Patent No. 3,773,319; EP 58,481), copolymers of L-glutamic acid and gamma ethyl-L-glutamate (Sidman et al., Biopolymers, 22, pp. 547-56 (1985)); poly(2-hydroxyethyl-methacrylate) or ethylene vinyl acetate (Langer et al., J. Biomed. Mater. Res., 15, pp. 167-277 (1981); Langer, Chem. Tech., 12, pp. 98-105 (1982)).

Liposomes containing inhibitors of CAAX processing enzymes may be prepared by well-known methods (See, e.g. DE 3,218,121; Epstein et al., Proc. Natl. Acad. Sci. U.S.A., 82, pp. 3688-92 (1985); Hwang et al., Proc. Natl. Acad. Sci. U.S.A., 77, pp. 4030-34 (1980); U.S. Patent Nos. 4,485,045 and 4,544,545). Ordinarily the liposomes are of the small (about 200-800 Angstroms) unilamellar type in which the lipid content is greater than about 30 mol.% cholesterol. The proportion of cholesterol is selected to control the optimal rate of MAG derivative and inhibitor release.

The inhibitors of mammalian CAAX processing

enzymes of this invention may also be attached to liposomes, which may optionally contain other agents to aid in targeting or administration of the compositions to the desired treatment site. Attachment of inhibitors to liposomes may be accomplished by any known cross-linking agent such as heterobifunctional cross-linking agents that have been widely used to couple toxins or chemotherapeutic agents to antibodies for targeted delivery. Conjugation to liposomes can also be accomplished using the carbohydrate-directed cross-linking reagent 4-(4-maleimidophenyl) butyric acid hydrazide (MPBH) (Duzgunes et al., J. Cell. Biochem. Abst. Suppl. 16E 77 (1992)).

The following are examples which illustrate the invention. These examples should not be construed as limiting: these examples are included for purposes of illustration only.

Example 1: Cloning and Sequencing of the Human *hRCE* cDNA

Isolation of 13R/27R λgt11 clones

To isolate the human cDNA counterpart to the yeast *RCE1* gene, we screened a λgt11 cDNA library made from a 9-week old human fetus. The library (ATCC #77433) was purchased from the American Type Culture Collection. Approximately 6 x 10^5 phage were plated onto the host E. coli strain LE392 and lifted onto Biodyne nylon membranes (ICN) by standard methodology (see, e.g., Sambrook et al., Molecular Cloning -- A Laboratory Manual [2nd ed.] Vol. 1-3 (1989)).

A plasmid vector containing the human EST (ID#358628) with similarity to the yeast Rce1p protein was purchased from Research Genetics and served as a probe. The EST was recovered from the vector by digesting with *Eco*RI restriction enzyme and agarose gel purification.

The EST, approximately 750 base pairs in length, was labeled with ^{32}P by random priming and hybridized to the phage lifts at high stringency according to the nylon membrane manufacturers specifications at 65°C for 18 hours. The hybridization buffer consisted of 5X Denhardt's solution, 5X SSC, 0.1% SDS, and 100 µg/ml denatured salmon sperm DNA. The filters were washed at high stringency at 65°C in 0.1% SDS and 0.1% SSC and exposed to autoradiography.

Two plaques, 13R and 27R, which showed a strong hybridization signal to the human EST (#358628) were purified to homogeneity by two successive rounds of plating and hybridization according to standard procedures.

Phage stocks were prepared by overlaying plates of *E. coli* strain LE392 that had been infected with λgt11 clones 13R and 27R with 20 mls SM. After gentle shaking at room temperature, approximately 12 mls was drawn off each plate, the cells were pelleted, and the stocks were stored at 4°C with a few drops of chloroform. The stocks were titered with LE392 and determined to be $2.6x10^3$ pfu/µl (13R) and $1.6x10^3$ pfu/µl (27R).

To prepare DNA from these two clones, LE392 cells (fresh overnight culture concentrated in 40% volume in SM) were infected with 100,000 pfu of 13R and 27R. The infection was incubated for 30 minutes at 42°C, plated on large LB plates with top agar, and incubated at 42°C overnight. The plates were then overlayed with 15 mls λ diluent and allowed to shake gently for 2 hours at room temperature. The λ diluent was then drawn off and the cells pelleted out. DNA was then prepared following standard procedures (see, e.g., Sambrook et al., Molecular Cloning -- A Laboratory Manual [2nd ed.] Vol. 1-3 (1989)), and resuspended in TE.

PCR Amplification

The 5′ ends of the λ clones were PCR amplified using two different internal primers (608RI, 5′-GGTTGCGTACAGACACTCCGTC-3′(SEQ ID NO:39); RCE 1R-RI, 5′-TTATTATACCACTTTTACCCG-3′(SEQ ID NO:40)) and an external λ primer (OMS178, 5′-TGGCGCAGGTAGCAGAG-3′(SEQ ID NO:41)) using Vent DNA polymerase). The PCR reactions were then treated with Taq DNA polymerase for 10 minutes at 72°C to produce A overhangs. The reactions were extracted with phenol/chloroform, precipitated with NaOAc/EtOH and resuspended in TE. The Invitrogen® Original TA Cloning® kit was used to clone the PCR products into pCR® 2.1 and transformed into INVαF′. Five clones containing inserts were identified: 13R/608-178-1, -3, 13R/RCE-178-3, 27R/608-178-1, -3. These five clones were sequenced using the ABI 377 sequencer.

The 3′ ends of the λ clones were subcloned by PCR amplifying the ends with an internal primer (8F, 5′-CTGCCTCACAGACATGCGTTG-3′(SEQ ID NO:42)) and an external λ primer (OMS179, 5′-GCGAAATACGGGCAGACATGG-3′(SEQ ID NO:43)) using Taq DNA polymerase. The Invitrogen® Original TA Cloning® kit was used to clone the PCR products into pCR® 2.1 and transformed into INVαF′. Three clones containing inserts were identified: 13R/8F-179-1, 27R/8F-179-1, -2. These three clones were sequenced using the ABI 377 sequencer.

Additional sequence information for the human *hRCE* cDNA was obtained by 5′ and 3′ RACE methods (Siebert et al., Nucleic Acids Research, 23, pp. 1087-1088 (1995)). Adaptor-ligated cDNA constructed from total human fetus mRNA (#7438-1), human placenta (#7411-1) and murine 15-day embryo (#7459-1) were purchased from Clonetech (Palo Alto, CA). Reactions were carried out under the recommended conditions of the manufacturer which included the following PCR conditions: 1X KlenTaq PCR buffer (Clonetech), 200μM dNTPs, 200nM oligonucleotide primers, 10 pg/μl cDNA, and 5 units of KlenTaq polymerase in a 50μl

reaction volume in the presence or absence of 5% DMSO. All PCR reactions included the primer AP1 (Clonetech) which is complementary to the cDNA adapter sequence and a gene specific primer. PCR reactions with human cDNA included one of the gene specific primers hRCE(1R)(SEQ ID NO:7), hRCE(2R)(SEQ ID NO:8), hRCE(10R)(SEQ ID NO:16) or hRCE(15R)(SEQ ID NO:21) for 5' RACE reactions and hRCE(3F)(SEQ ID NO:9) or hRCE(4F)(SEQ ID NO:10) for 3' RACE reactions. 5' RACE reactions of murine cDNA included AP1 (Clonetech) and one of the following gene specific primers: 608(1R)(SEQ ID NO:27), 608(2R)(SEQ ID NO:28) or hRCE(10R)(SEQ ID NO:16). All PCR reactions were performed in an ABI 9600 thermocycler. 5' and 3' RACE products were resolved by agarose gel electrophoresis, excised and recovered. The DNA products were subcloned into the T/A cloning vector, pCR2.1 (Invitrogen, Carlsbad, CA) and sequenced.

Example 2: Human hRCE RNA Analysis

Northern Blot Analysis

Two multiple tissue mRNA blots (#7760-1, #7759-1) were purchased from Clonetech and hybridized to the human EST (ID#358628) uniformly labeled with ^{32}P as described in **Example 1**. Hybridization and washing were done at high stringency at 65°C using the conditions recommended by Clonetech.

Primer Extension Analysis

5 µg human poly(A) RNA from pancreas (Clonetech, catalog # 6539-1) was hybridized to 0.9 pmol of ^{32}P end-labeled primer (15R, 5'-AGGCTGCCCACGTAGGAGCAGGCGAGGCTGA-3' (SEQ ID NO:21) $2.6x10^6$ cpm/pmol) in hybridization buffer (2 mM Tris [pH 7.6], 0.2 mM EDTA, 250 mM KCl) in a volume of 10 µl. The reaction mixture was placed at 73°C for 2 minutes and then at 53°C for 1 hour. Extension of the

primer was initiated by the addition of 40 µl of reverse transcriptase buffer (20 mM Tris [pH 8.7], 10 mM $MgCl_2$, 100 µg/ml actinomycin D, 5 mM dithiothreitol, 0.33 mM deoxynucleoside triphosphates) and 200 units SuperScript™II (GibcoBRL, catalog # 18064-014). The extension reaction was placed at 37°C for 30 minutes and then terminated with the addition of 300 µl ethanol. The nucleic acids were pelleted and resuspended in 5 µl formamide loading buffer/0.1 M NaOH (2:1). The reaction products were then separated on an 8% polyacrylamide/urea sequencing gel, and visualized by autoradiography. The predominant extension product was approximately 41 nucleotides long.

To control for the possibility of secondary structures at the 5' end of the *hRCE1* mRNA affecting the primer extension analysis, the reactions were performed utilizing the GeneAmp® Thermostable r*Tth* Reverse Transcriptase RNA PCR kit (Perkin Elmer, catalog # N808-0069). Human Poly(A) RNA (5 µg) from pancreas (Clontech, catalog # 6539-1) and placenta (Clontech, catalog # 6518-1) were separately hybridized to 0.9 pmol of ^{32}P end-labeled primer (15R, $2.6x10^6$ cpm/pmol (SEQ ID NO:21)) in r*Tth* reverse transcriptase buffer (10 mM Tris [pH 8.3], 90 mM KCl) in a volume of 14.4 µl. The hybridization was carried out at 70°C for 15 minutes. 5.6 µl extension mix (final concentrations: 1 mM $MnCl_2$, 1 mM dATP, 1 mM dCTP, 1 mM dGTP, 1 mM dTTP, 5 units r*Tth* DNA polymerase) was added to the hybridization reactions and incubated at 70°C for 20 minutes. The reactions were terminated with the addition of 120 µl ethanol and the nucleic acids pelleted. The products were resuspended in 5 µl formamide loading buffer/0.1 M NaOH (2:1). In order to size the primer extension products double stranded sequencing reactions were performed using 15R (SEQ ID NO:21) as the primer and a subclone of the λ clone 27R (27R/608-178-1, see above).

The sequencing reactions were performed utilizing the Sequenase Version 2.0 DNA Sequencing Kit (United States Biochemical, catalog # US70770) as per the kit instructions. The primer extension products and the sequencing reactions were resolved on an 8% polyacrylamide/urea sequencing gel followed by autoradiography.

Example 3: Expression of the human *hRCE* cDNA in a yeast mutant

The ORF beginning at position 116 of the *hRCE* gene (SEQ ID NO:1, nucleotides 116 to 997; **Figure 2A**) was expressed in a yeast strain deleted for both CAAX protease genes, y*AFC1* and y*RCE1* (Boyartchuk et al., supra) (**Example 3**). To create the expression construct, we amplified by PCR the human *hRCE* ORF from the first valine (defined by nucleotides 116-118) to the stop codon (defined by nucleotides 995-997) employing the human *hRCE* cDNA λ clone 27R as a template. The primers included in the reaction were hRCE-ValX (SEQ ID NO:22) and hRCE3P-RI (SEQ ID NO:23). The primer hRCE-ValX incorporated an *Xba*I restriction site immediately upstream (5') of the GTG valine codon (defined by SEQ ID NO:1, nucleotides 116-118). The primer hRCE3P-RI incorporated an *Eco*RI site immediately downstream (3') of the TGA stop codon (defined by SEQ ID NO:1, nucleotides 995-997).

Following amplification, the PCR product was digested with *Eco*RI and *Xba*I and subcloned into the respective sites of the expression vector, pACA1' or pACA5'. pACA1' contains 1000 base-pairs of the yeast *MEV1* promoter including the translational initiation codon ATG. The *MEV1* promoter was amplified by PCR from yeast genomic DNA using the following primers:

MEV1-1 (containing *Hin*dIII site): 5'-GGA AAA GCT TTT GAA GCA CCT GTG GAG T-3' (SEQ ID NO:44); and

MEV1-2 (containing *Xba*I site): 5'-GTA AAC GTT CTA GAC ATT TTG AGT ACG TC-3' (SEQ ID NO:45).

Vector pACA5′ contains 700 base-pairs of the yeast *ERG12* promoter and the first four codons of the *ERG12* coding sequence. The *ERG12* promoter was amplified by PCR from yeast genomic DNA using the following primers:

ERG12-1 (containing *Sph*I site): 5'-GGC CGG GCA TGC AAC CTA TTC GAG ATG AAT GCG-3' (SEQ ID NO:46); and

ERG12-2 (containing *Xba*I site): 5'-CCG GCC TCT AGA CGG TAA TGA CAT ATT GAC AGT-3' (SEQ ID NO:47).

Both expression vectors include an *Eco*RI-*Nar*I fragment containing the yeast PGK1 transcriptional termination sequence. A fragment of DNA comprising the yeast PGK1 termination sequence was amplified by PCR from yeast genomic DNA using the following primers:

PGK 1 (containing an EcoRI site): 5'-GAT TGA ATT CAA TTG AAA TCG ATA G-3' (SEQ ID NO:48); and

PGK 2 (containing a NarI site): 5'-CCG AGG CGC CGA ATT TTC GAG TTA T-3' (SEQ ID NO:49).

The resulting plasmids, pACA7001 (alternatively referred to herein as pAB14) and pACA7002 (alternatively referred to herein as pAB13), represent translational fusions between the *MEV*1 ORF and the *ERG*12 ORF, respectively.

The yeast strain JRY5317 (Boyartchuk et al., supra) is an a-mating type strain in which both CAAX protease genes have been deleted. For transformation of the *hRCE* expression plasmids, JRY5317 was grown overnight in rich medium (2% Bacto Peptone, 1% Yeast Extract, 2% Glucose). The transformation was performed by the lithium acetate method (Gietz and Schiestl, Yeast, 7, pp. 253-263 (1991)). Transformants were selected for uracil prototrophy on minimal medium.

Plasmids pACA7001 and pACA7002 were each transformed into the yeast strain JRY5317, an a-mating

type strain in which both CAAX protease genes have been deleted (Boyartchuk et al., supra)(**Example 3**). Halo assays were performed by replica plating patches of the yeast transformants onto a lawn of α *sst2* tester lawn as described (Boyartchuk et al., supra). Cells harboring an *sst2* mutation are rendered supersensitive to mating pheromone and undergo a G1 arrest when exposed to pheromone. This creates a halo around the patch of mated yeast cells.

Example 4: Expression of the human *hAFC* cDNA in a yeast mutant.

The ORF described herein which encodes a homolog of the yeast *AFC1* gene (SEQ ID NO:37; **Figure 8**) is expressed in a yeast strain deleted for both CAAX protease genes, y*AFC1* and y*RCE1* (Boyartchuk et al., supra) as described in **Example 3**. To create the expression construct, the human *AFC* ORF (SEQ ID NO:37) is amplified from the first GCG valine codon to the stop codon at ca. nucleotide 721 (see **Figure 7**) and inserted into yeast expression vectors as described in **Example 3**.

Example 5: A screening method for identifying inhibitors of mammalian CAAX processing enzymes.

A surrogate organism that is easy to manipulate and maintain, such as *Saccharomyces cerevisiae*, may be transformed with an expression vector encoding *hRCE* or *hAFC* sequences. Such a recombinant organism may serve as a template for *in vivo* screening for inhibitors of the protease.

Such an assay could be direct or indirect. A direct assay could be created based upon the cleavage of a CAAX-containing proteins such as a-factor in a halo assay or mating assay as described in (Boyartchuk et al.,

supra). Another assay could be based upon the requirement of CAAX proteolysis for Ras function. Such an assay would utilize a yeast strain possessing *hRCE* and an activated allele of Ras such as the Gly19 to Val19 mutation. Cells that harbor an activated allele of Ras are susceptible to acute heat shock at 55°C. Thus, one could screen for inhibitors of Ras by screening or selecting compounds that cause a heat-shock resistant phenotype.

Alternatively, an assay for hRCE could be based upon an indirect effect of modulating the CAAX protease activity. For example, cells could be transformed with either a normal or non-functional version of *hRCE*. The expression of other selected genes in an organism could then be assessed. Once a gene was found whose expression was dependent upon the function of *hRCE*, that gene could then serve as the basis of an assay.

Example 6: A screening method for identifying proteins that interact with mammalian CAAX processing enzymes

A protein which physically interacts with a mammalian CAAX protease according to this invention may be isolated using the "two-hybrid system" (see, e.g., C.T. Chien et al., "The two-hybrid system: a method to identify and clone genes for proteins that interact with a protein of interest", Proc. Natl. Acad. Sci. U.S.A., 88(21), pp. 9578-82 (1991)). This method takes advantage of the multi-domain structure of the yeast transcriptional activator protein GAL4, which has a DNA binding domain separable from a transcriptional activation domain. Expression plasmids are constructed to encode one of two hybrid proteins. One hybrid consists of the GAL4 DNA-binding domain fused to a mammalian CAAX protease polypeptide of this invention. The second hybrid consists of the GAL4 activation domain fused to protein sequences

encoded by a library of genomic DNA or cDNA fragments from a mammalian or other cellular source. (The plasmids are co-transformed into an appropriate yeast strain, or are individually transformed into yeast strains of opposite mating types having genetic markers that allow selection of diploids.) Interaction between the CAAX protease and a protein encoded by one of the library plasmids leads to transcriptional activation of a reporter gene of choice which contains an appropriately positioned GAL4 binding site. The advantage of this method is that the gene which encodes the protein that interacts with the CAAX protease may be identified directly from the library.

Example 7: Mutagenesis of the Histidine-Rich Region of *hRCE1*

Each histidine comprising the histidine-rich sequence "HxHH" of the human Rce1 protein (SEQ ID NO:2, amino acids 172-175) were changed individually to an alanine residue by site-directed mutagenesis.

Plasmids pAB15, pAB16, and pAB17 were constructed as follows: An 888 bp XbaI-EcoRI fragment was cut from pAB13 (see pACA7002, **Example** 3) and cloned into pBluescript SK. Histidines 172, 174 and 175 were each mutated to an alanine using the QuikChange™ Site-Directed Mutagenesis Kit following manufacturer's directions (Stratagene). The following complementary pairs of oligonucleotides were used (underlined bases are those that were changed relative to the wild-type sequence (5'-CTC TTT TTT GGA GTT GCC CAT TTT CAC CAT ATT ATT GAG CAG CTG -3') (SEQ ID NO:1, nucleotides 611-655):

<u>For H172A</u>

hRCE1-M1: 5'-CTC TTT TTT GGA GTT GCC <u>GCT</u> TTT CAC CAT ATT ATT GAG CAG CTG-3'

hRCE1-M2: 5'-CAG CTG CTC AAT AAT ATG GTG AAA AGC GGC AAC TCC AAA AAA GAG-3'

For H174A

hRCE1-M3: 5'-CTC TTT TTT GGA GTT GCC CAT TTT GCC CAT ATT ATT GAG CAG CTG-3'

hRCE1-M4: 5'-CAG CTG CTC AAT AAT ATG GGC AAA ATG GGC AAC TCC AAA AAA GAG-3'

For H175A

hRCE1-M5: 5'-GGA GTT GCC CAT TTT CAC GCT ATT ATT GAG CAG CTG CGT TTC-3'

hRCE1-M6: 5'-GAA ACG CAG CTG CTC AAT AAT AGC GTG AAA ATG GGC AAC TCC-3'

Mutations were confirmed by sequencing on an ABI Prism 377 DNA Sequencer using the ABI Prism Dye Terminator Sequencing Kit with AmpliTaq DNA Polymerase, FS. The 888 bp XbaI-EcoRI fragments from those clones with the desired mutations were then cloned back into pAB13 (replacing the corresponding wild-type *hRCE1* XbaI-EcoRI fragment) generating pAB15 (hRCE1-H172A), pAB16 (hRCE1-H174A) and pAB17 (hRCE1-H175A).

Example 8: Whole Cell Co-Culture Assay for Identifying Rce1p Inhibitors

A strain of yeast is constructed in which the endogenous *RAS1* gene has been deleted and the endogenous *RAS2* gene has been replaced with the temperature-sensitive *ras2-23* allele, which sensitizes the strains such that their growth rates depend upon the function of Rce1p. Cells from this strain are transformed with a high-copy (2μ) plasmid (e.g., a LEU2, URA3 or HIS3-marked plasmid

whose presence can be selected in the appropriate leu, ura or his auxotrophic strain background) which either does (Strain 1) or does not (Strain 2) additionally encode: a) the *RCE1* gene driven by a strong (e.g., the *ADH1*) promoter; and b) a marker gene (e.g., green fluorescent protein (GFP)) under the control of a strong promoter. Strain 1 and Strain 2 transformants are selected on the appropriate medium (e.g., selecting for leu, ura or his prototrophy). Strain 1 (but not Strain 2) transformants are also selected based on their fluorescence (from expression of the plasmid-linked GFP marker).

Strain 1 and Strain 2 are grown independently to a point where they are in mid-log phase ($O.D._{600}$ = 1.0) and are then mixed at an initial ratio of 1 cell of Strain 1 for every 5,000 cells of Strain 2. The mixture does not exhibit significant fluorescence because Strain 1, which harbors the multi-copy plasmid encoding *RCE1* and the fluorescent marker GFP, is present in such low abundance. Following the initial mixture of the two strains, the co-culture is aliquoted into wells of 96-well plates.

A typical 96-well plate will handle about 0.25ml of liquid. One OD_{600} of yeast in rich medium (YPD) is about 1×10^7 cells/ml. Taking into consideration that we want a 5,000-fold excess of Strain 2 in the mixture and we want to have at least 50 cells of Strain 1 in the mixture to take into account inevitable variations, then a co-culture experiment where the starting OD_{600} = 0.1 would include 50 cells of Strain 1 and 250,000 cells of Strain 2. This gives a Strain 1-to-Strain 2 ratio of 1:5,000. A typical yeast overnight growth in a 96-well plate will produce an OD_{600} ~4. This translates to a 40-fold increase in growth, or 5.3 generations.

To identify competitive inhibitors of Rce1p, a dilution series (e.g., ten-fold) of each selected potential inhibitor resuspended in an appropriate medium

is prepared and an aliquot of each test sample added individually to the wells of the 96-well plates which contain the co-cultured strains. As a control, an equal volume of each medium in which the test inhibitors are suspended (now without the test inhibitor) is added to a well containing the strain co-culture. The co-cultures plus test chemicals and controls are incubated until the cells have divided numerous times. The cells are diluted into fresh medium and allowed to undergo ca. 5 generation doublings (as determined by spectrographic absorbance readings). At various points during this growth period, the relative abundance of Strain 1 in the mixture is assessed by fluorescence emission from the 96-well plates.

Example 9: Heat Shock Assays for Measuring hRCE Function and for Identifying Rcelp Inhibitors

Plasmids

Plasmid pAB1 was constructed as follows: The backbone plasmid for pAB1 was Yeplac195 (2μ, *URA3*; Gietz and Sugino, 1988) and contains an insert consisting of two parts: green fluorescent protein (GFP) and phosphoglycerate kinase (*PGK1*) terminator. The coding region of GFP from *Aequoria victoria* containing the S65T mutation (Ormo et al., 1996) was amplified by PCR using oligonucleotides:

5P-GFP-ORF (5'-CAT GTC TAG AGG AGA AGA ACT TTT C-3') (SEQ ID NO:51); and

3P-GFP-ORF (5'-CGC GAA TTC CTA TTT GTA TAG TTC A-3'); and has XbaI and EcoRI sites engineered on the 5' and 3' ends, respectively.

The *PGK1* terminator was amplified by PCR using the following oligonucleotides:

PGK 1 (containing an EcoRI site): 5'-GAT TGA ATT CAA

TTG AAA TCG ATA G-3' (SEQ ID NO:48); and

PGK 2 (containing a NarI site): 5'-CCG AGG CGC CGA ATT TTC GAG TTA T-3' (SEQ ID NO:49) and consists of 263 bp immediately downstream of the stop codon of *PGK1*. EcoRI and NarI sites were engineered on the 5' and 3' ends of the *PGK1* terminator.

Plasmid pAB20 was constructed as follows: The dominant allele of *RAS2*, *RAS2*^{val19} (D. Broek, et al., Cell, 41, pp. 763-769 (1985)), was cloned as a XhoI-HindIII fragment (1906 bp) into the *CEN-LEU2* plasmid pRS315 (R.S. Sikorski and P. Hieter, Genetics, 122, pp. 19-27 (1989)). This plasmid carries the coding region for Ras2p^{val19} as well as 639 bp upstream and 298 bp downstream of the *RAS2* open reading frame.

Plasmids pAB13 *hRCE1*, pAB15 H172A, pAB16 H174A; and pAB17 H175A were made as described in **Examples 3** and **7**.

Heat shock assays in yeast cells

Yeast cells wild-type for *RCE1* (W303) (*MATa; leu2-3,112; ade2-101; his3-11; trp1-1; ura3-1*) were used to create an isogenic, double mutant strain lacking both the Rce1p and Afc1p CAAX proteases (*Δrce1, Δafc1*). The wild-type (W303) or *Δrce1, Δafc1* strains were separately transformed with plasmid pAB20 expressing the dominant mutant RAS2^{val19} gene. Transformed *Δrce1, Δafc1* cells then were co-transformed with one of the following *URA3* marked, high-copy (2μ) plasmids: pAB1 (vector control); pAB13 (*hRCE1*), pAB15 (H172A), pAB16 (H174A); and pAB17 (H175A) (as described in **Figure 10**).

Yeast cell transformants were grown at 30°C for approximately 48 hours in synthetic complete media lacking uracil and leucine and supplemented with 2% glucose, to an approximate OD600 of 4-6. The cultures were diluted 1:10

in water in 0.5 ml PCR tubes and placed in a PTC-100 thermocycler (MJ Research, Inc.). The heat-shock regimen was performed as follows: 25°C 3 min; 55°C 1 min, 25°C 3 min. The cultures were further diluted in water and plated for viability. The number of surviving colonies were compared with those from parallel samples that had not undergone heat shock. The results of this experiment are shown in **Figure 11.**

Heat shock assays to identify CAAX protease inhibitors

The above heat shock assay may be modified by one of skill in the art to identify inhibitors of Rce1p in the presence of an activated *RAS2* allele. A dilution series of each selected potential inhibitor may be tested (e.g., as described in **Example 8**). Rce1p inhibitors may be identified by their ability to increase in a concentration-dependent manner the viability of heat-shocked cells which express activated Ras2p, which express Rce1p from a chromosomal and/or from a single or multi-copy plasmid. In addition, the cells can be deleted for the *yRCE1* gene to identify inhibitors of the human hRce1p protein.

What is claimed is:

1. An isolated, recombinant or synthetic DNA molecule which comprises a nucleic acid sequence selected from SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5 or SEQ ID NO: 37.

2. The DNA according to claim 1 wherein the DNA comprises a nucleic acid sequence of SEQ ID NO: 1.

3. An isolated, recombinant or synthetic DNA or polynucleotide comprising a nucleic acid sequence that is substantially homologous to any of the DNA molecules of claims 1 or 2.

4. An isolated, recombinant or synthetic DNA or polynucleotide comprising a nucleic acid sequence that hybridizes under stringent conditions to any of the nucleic acid sequences of claims 1 or 2.

5. The DNA or polynucleotide according to any one of claims 3 or 4, wherein said polynucleotide is selected from the group consisting of RNA, cDNA, genomic DNA, synthetic nucleic acids, chemically or biochemically modified nucleic acid comprising non-natural or derivatized nucleic acid bases, or mixed polymers of any of the above.

6. An isolated, recombinant or synthetic DNA or polynucleotide comprising the antisense RNA or antisense DNA sequence of the DNA or polynucleotide according to any one of claims 1-5.

7. A DNA molecule according to any one of claims 1-6, further comprising an expression control sequence operatively linked to the DNA or

polynucleotide sequence.

8. A polynucleotide comprising at least eight consecutive nucleotides of the DNAs of claims 1 or 2 or the polynucleotides of claims 3-6.

9. A vector comprising a DNA according to claim 1 or 2 or a polynucleotide according to claims 3-6.

10. A host cell transformed with a molecule according to any one of claims 1 to 9.

11. The host cell according to claim 10, wherein the molecule is a DNA molecule which is integrated into the genome of said host cell.

12. A method for producing a polypeptide comprising the step of culturing a host cell transformed with a DNA molecule according to claim 7.

13. A polypeptide encoded by a DNA or polynucleotide comprising a nucleic acid sequence selected from SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5 or SEQ ID NO: 37.

14. A polypeptide encoded by a DNA or polynucleotide comprising a nucleic acid sequence selected from the group consisting of:

(a) a nucleic acid sequence that is substantially homologous to any of the nucleic acid sequences of claims 1-5; and

(b) a nucleic acid sequence that hybridizes under stringent conditons to any one of the nucleic acid sequences of claims 1-5.

15. A polypeptide comprising amino acid sequences

selected from SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6 or SEQ ID NO: 38.

16. A polypeptide characterized by being at least 80% identical in amino acid sequence to any one of the amino acid sequences selected from SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6 or SEQ ID NO: 38.

17. An antibody that binds specifically to an epitope of any one of the polypeptides of claims 13-15.

18. A pharmaceutical composition comprising any one of the polypeptides of claims 13-15.

19. A pharmaceutical composition comprising an inhibitor of a mammalian CAAX processing enzyme, said enzyme comprising amino acid sequences selected from SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6 or SEQ ID NO: 38.

20. A method for identifying a competitive inhibitor of any one of the polypeptides of claims 13-15, wherein said method comprises the steps of:

a) co-culturing two distinguishable strains of cells which differ by the presence or absence of a functional *RCE1* or *AFC1* gene; and

b) selecting compounds which increase the relative proportion of cells expressing the functional *RCE1* or *AFC1* gene compared to those which do not.

21. A method for treating a CAAX-protein mediated disease or disorder in a patient by administering a pharmaceutical composition comprising any one of the

polypeptides of claims 13-15.

22. A method for treating a CAAX-protein mediated disease or disorder in a patient by administering a pharmaceutical composition according to claim 19.

NCBI XREFdb ? ?

Cross-referencing the Genetics of Model Organisms with Mammalian Phenotypes

Results for TBLASTN Search vs. dbEST:

Gene name(s): RCE1 Searched: 26 May 97

NCBI ID:	GenBank Accession:	Org:	P value:	N:	Hiscore:	Mapped (XREF)	Mapped (Other)	New?
604642	W96411	Human	1.4e-19	4	79	No	No	

Clone ID: 358628
Organism: Homo Sapiens
Library: Soares fetal heart NbHH19W
Source: Wilson RK
E-mail: est@watson.wustl.edu
Phone: 314 286 1800 Fax: 314 286 1810

To order this IMAGE Consortium clone, visit either the Genome Systems, Research Genetics, or ATCC World Wide Web pages.

Minus Strand HSPs

```
 Score = 79 (37.3 bits), Expect = 1.4e-19, Sum p(4) = 1.4e-19
 Identities = 14/28 (50%), Positives = 19/28 (67%), Frame = -3

Query:   211 SILLTTCFQILYTTLFGGLTKFVFVRTG 238
             +I L+  FQ  YT +FG  T F+F+RTG
Sbjct:   162 NIFLSAAFQFSYTAVFGAYTAFLFIRTG 79

 Score = 61 (28.8 bits), Expect = 1.4e-19, Sum P(4) = 1.4e-19
 Identities = 11/15 (73%), Positives = 12/15 (80%), Frame = -1

Query:   188 LFFGLAHAHHAYEQL 202
             LFFG+AH HH  EQL
Sbjct:   230 LFFGVAHFHHIIEQL 186

 Score = 53 (25.0 bits), Expect = 1.4e-19, Sum P(4) = 1.4e-19
 Identities = 9/19 (47%), Positives = 13/19 (68%), Frame = -3

Query:   239 GNLWCCIILHALCNIMGFP 257
             G+L   ++ H+ CN MGFP
Sbjct:    81 GHLIGPVLCHSFCNYMGFP 25

Score = 48 (22.7 bits), Expect = 1.4e-19, Sum P(4) = 1.4e-19
 Identities = 8/19 (42%), Positives = 14/19 (73%), Frame = -1

Query:   147 RNFIFAPITEEIFYTSMLL 165
             RN + AP+TEE+ + + +L
Sbjct:   338 RNQVIAPLTEELVFRACML 282
```

FIG. 1A

NCBI XREFdb ? ?

Cross-referencing the Genetics of Model Organisms with Mammalian Phenotypes

Results for TBLASTN Search vs. dbEST:

Gene name(s): RCE1 Searched: 26 May 97

NCBI ID:	GenBank Accession:	Org:	P value:	N:	Hiscore:	Mapped (XREF)	Mapped (Other)	New?
482520	N76181	Human	8.4e-15	1	168	No	No	

Clone ID: 284572
Organism: Homo Sapiens
Library: Soares multiple sclerosis 2NbHMSP
Source: Wilson RK
E-mail: est@watson.wustl.edu
Phone: 314 286 1800 Fax: 314 286 1810

To order this IMAGE Consortium clone, visit either the Genome Systems, Research Genetics, or ATCC World Wide Web pages.

Plus Strand HSPs

Score = 168 (77.8 bits), Expect = 8.4e-15, P = 8.4e-15
Identities = 36/80 (45%), Positives = 51/80 (63%), Frame = +2

```
Query: 119 MLFNDLLTPLECAMQFVMSLISRTHEYQADAYAKKLGYKQNLCRALIDLQIKNLSTMNVD 178
           ++F  + +P    + F ++++SR  E+QADA+AKKLG  ++L  ALI L   NL     D
Sbjct: 32  IIFQFIFSPYNEVLSFCLTVLSRRFEFQADAFAKKLGKAKDLYSALIKLNKDNLGFPVSD 211

Query: 179 PLYSSYHYSHPTLAERLTAL 198
            L+S +HYSHP L ERL AL
Sbjct: 212 WLFSMWHYSHPPLLERLQAL 271
```

A human, mouse, and rat subset of dbEST is used for all XREFdb EST searching. Default BLAST parameters and SEG low complexity masking are used for each Search.

Go Back to the complete hit list for this query.

Return to your Query Summary.

Online Help is available for this form, simply click any label for assistance.

Xref info@gmail.bs.jhu.edu
Douglass E. Bassett, Jr. (bassett@ncbi.nlm.nih.gov)
http://www.ncbi.nlm.nih.gov/XREFdb/

FIG. 1B

NCBI XREFdb ?

Cross-referencing the Genetics of Model Organisms with Mammalian Phenotypes

Results for TBLASTN Search vs. dbEST:

Gene name(s): RCE1 Searched: 26 May 97

NCBI ID:	GenBank Accession:	Org:	P value:	N:	Hiscore:	Mapped (XREF)	Mapped (Other)	New?
636296	AA022288	Mouse	1.6e-8	1	126	Ordered	No	NEW

Clone ID: 456971
Organism: Mus musculus
Library: Soares mouse placenta 4NbMP13.5 14.5
Source: Marra M/Mouse EST Project
E-mail: mouseest@watson.wustl.edu
Phone: 314 286 1800 Fax: 314 286 1810

To order this IMAGE Consortium clone, visit either the Genome Systems, Research Genetics, or ATCC World Wide Web pages.

Plus Strand HSPs

Score = 126 (59.2 bits), Expect = 1.6e-08, P = 1.6e-08
Identities = 25/68 (36%), Positives = 41/68 (60%), Frame = +1

```
Query:  43 QIRFKNFNLFIIALSLFHFLEYYITAKYNPLKVHSESFLLNNGKSYMAAHSFAILECLVE 102
           Q  + +F  ++ +LSLFH+ EY +TA  NP  +  +SFLLN+   y  A   + +E  +E
Sbjct: 316 QSSWNHFGWYVCSLSLFHYSEYLVTAVNNPKSLSLDSFLLNHSLEYTVAALSSWIEFTLE 495

Query: 103 SFLFPDLK 110
           +  +P+LK
Sbjct: 212 NIFWPELK 519
```

This EST clone was **selected for mapping** by the XREF project based upon Similarity to the STE14/YDR410C/D9461.1 gene (TBLASTN P-value = 1.6e-8).

Order authorized: 15 Jan 97
Clone ordered: 18 Apr 97
Clone received: 24 Apr 97

A human, mouse, and rat subset of dbEST is used for all XREFdb EST searching. Default BLAST parameters and SEG low complexity masking are used for each search.

Go Back to the complete hit list for this query.

Return to your Query Summary.

Online Help is available for this form, simply click any label for assistance.

FIG. 1C

```
GAATTCCGGGGCGGCGCTGGGCGGGGATGGGCTGCGACTGCTGICGGIATCGCGCGGCCGGAGCGGCCGCCCGAGTCGGCGGCGGCTGGGCGGCCTGGGCCCC
CTTAAGGCCCCGCCGCGACCCGCGCCCCTACCCGACGCTGACGACAGCCATAGCGCCGGCCTCGCCCGGCGGGCTCAGCCGCCGCGACCCGGCCGACCCGGGG
  N  S  G  A  A  L  G  G  D  G  L  R  L  L  S  V  S  R  P  E  R  P  P  E  S  A  A  L  G  G  L  G  G  P

GGGCTGTGTGCTGGGTGTCAGTGTTCTCCTGCCTCAGCCTCGCCTGCTCCTACGTGGGCAGCCTCTACGTCTGGAAGAGCGAACTGCCCAGGGACCATC
CCCGACACGACGACCCACAGTCACAAGAGGACGGAGTCGGAGCGGACGAGGATGCACCCGTCGGAGATGCAGACCTTCTCGCTTGACGGGTCCCTGGTAG
  G  L  C  C  W  V  S  V  F  S  C  L  S  L  A  C  S  Y  V  G  S  L  Y  V  W  K  S  E  L  P  R  D  H

CCGCGGTCATCAAGCGACGCTTCACCAGCGTCCTGGTGGTGTCCAGTCTCTCACCCCTGTGCGTGCTGCTCTGGAGGGAACTCACAGGCATCCAGCCAGG
GGCGCCAGTAGTTCGCTGCGAAGTGGTCGCAGGACCACCACAGGTCAGAGAGTGGGGACACGCACGACGAGACCTCCCTTGAGTGTCCGTAGGTCGGTCC
  P  A  V  I  K  R  R  F  T  S  V  L  V  V  S  S  L  S  P  L  C  V  L  L  W  R  E  L  T  G  I  O  P  G

CACATCCCTGCTCACCCTGATGGGCTTCAGGCTGGAGGGCATTTTCCCAGCGGCGCTGCTCCCCTGTTGCTGACCATGATTCTTTTCCTGGGCCCACTG
GTGTAGGGACGAGTGGGACTACCCGAAGTCCGACCTCCCGTAAAAGGGTCGCCGCGACGAGGGACAACGACTGGTACTAAGAAAAGGACCCGGGTGAC
  T  S  L  L  T  L  M  G  F  F  R  L  E  G  I  F  P  A  A  L  L  P  L  L  L  T  M  I  L  F  L  G  P  L

ATGCAGCTCTCTATGGATTGCCCTTGTGACCTGGCAGATGGGCTGAAGGTTGTCCTGGCCCCCCGCTCCTGGGCCCGCTGCCTCACAGACATGCGTTGGC
TACGTCGAGAGATACCTAACGGGAACACTGGACCGTCTACCCGACTTCCAACAGGACCGGGGGGCGAGGACCCGGGCGACGGAGTGTCTGTACGCAACCG
  M  O  L  S  M  D  C  P  C  D  L  A  D  G  L  K  V  V  L  A  P  R  S  W  A  R  C  L  T  D  M  R  W

TGCGGAACCAAGTGATCGCCCCGCTGACAGAGGAGCTGGTGTTCCGGGCCTGTATGCTGCCCATGTTAGCACCGTGCATGGGCCTGGGCCCTGCTGTGTT
ACGCCTTGGTTCACTAGCGGGGGCGACTGTCTCCTCGACCACACAAGGCCCCGGACATACGACGGGTACAATCGTGGCACGTACCCGGACCCGGGACGACACAA
  L  R  N  O  V  I  A  P  L  T  E  E  L  V  F  R  A  C  M  L  P  M  L  A  P  C  M  G  L  G  P  A  V  F
```

FIG. 2A-1

SUBSTITUTE SHEET (RULE 26)

```
CACCTGCCCGCTCTTTTTTGGAGTTGCCCATTTTCACCATATTATTGAGCAGCTGCGTTTCCGCCAGAGCAGCGTGGGGAACATCTTCTTGTCTGCTGCG
                                                                                                      700
GTGGACGGGCGAGAAAAAACCTCAACGGGTAAAAGTGGTATAATAACTCGTCGACGCAAAGGCGGTCTCGTCGCACCCCTTGTAGAAGAACAGACGACGC
  T  C  P  L  F  F  G  V  A  H  F  H  H  I  I  E  Q  L  R  F  R  Q  S  S  V  G  N  I  F  L  S  A  A

TTCCAGTTCTCCTACACAGCTGTCTTCGGTGCCTACACTGCTTTCCTCTTCATCCGCACAGGACACCTGATTGGGCCGGTTCTCTGCCATTCCTTCTGCA
                                                                                                      800
AAGGTCAAGAGGATGTGTCGACAGAAGCCACGGATGTGACGAAAGGAGAAGTAGGCGTGTCCTGTGGACTAACCCGGCCAAGAGACGGTAAGGAAGACGT
  F  Q  F  S  Y  T  A  V  F  G  A  Y  T  A  F  L  F  I  R  T  G  H  L  I  G  P  V  L  C  H  S  F  C

ATTACATGGGTTTCCCAGCTGTTTGCGCGGCCTTGGAGCACCCACAGAGGCGGCCCCTGCTGGCAGGCTATGCCCTGGGTGTGGGACTCTTCCTGCTTCT
                                                                                                      900
TAATGTACCCAAAGGGTCGACAAACGCGCCGGAACCTCGTGGGTGTCTCCGCCGGGGACGACCGTCCGATACGGGACCCACACCCTGAGAAGGACGAAGA
  N  Y  M  G  F  P  A  V  C  A  A  L  E  H  P  Q  R  R  P  L  L  A  G  Y  A  L  G  V  G  L  F  L  L  L

GCTCCAGCCCCTCACGGACCCCAAGCTCTACGGCAGCCTTCCCCTTTGTGTGCTTTTGGAGCGGGCAGGGGACTCAGAGGCTCCCCTGTGCTCCTGACCT
                                                                                                     1000
CGAGGTCGGGGAGTGCCTGGGGTTCGAGATGCCGTCGGAAGGGGAAACACACGAAAACCTCGCCCGTCCCCTGAGTCTCCGAGGGGACACGAGGACTGGA
  L  Q  P  L  T  D  P  K  L  Y  G  S  L  P  L  C  V  L  L  E  R  A  G  D  S  E  A  P  L  C  S  .  P

ATGCTCCTGGATACGCTATGAACTCTCACCGGCTCCCCAGCCCTCCCCCCAAGGGGTACTGCAGGGGAAGGGCTGGCTGGGGTCCCCGAGATCTCAGGAA
                                                                                                     1100
TACGAGGACCTATGCGATACTTGAGAGTGGCCGAGGGGTCGGGAGGGGGGTTCCCCATGACGTCCCCTTCCCGACCGACCCCAGGGGCTCTAGAGTCCTT
  M  L  L  D  T  L  .  T  L  T  G  S  P  A  L  P  P  R  G  T  A  G  E  G  L  A  G  V  P  E  I  S  G

TTTTTGTAGGGGATTGAAGCCAGAGCTAGTTGCGTCCCAGGGACCAAGAGAAAGAAGCAGATATCCAAAGGGTGCAGCCCCTTTTGAAAGGGGTGTTTAC
                                                                                                     1200
AAAAACATCCCCTAACTTCGGTCTCGATCAACGCAGGGTCCCTGGTTCTCTTTCTTCGTCTATAGGTTTCCCACGTCGGGGAAAACTTTCCCCACAAATG
  I  F  V  G  D  .  S  O  S  .  L  R  P  R  D  Q  E  K  E  A  D  I  Q  R  V  Q  P  L  L  K  G  V  F  T
```

FIG. 2A-2

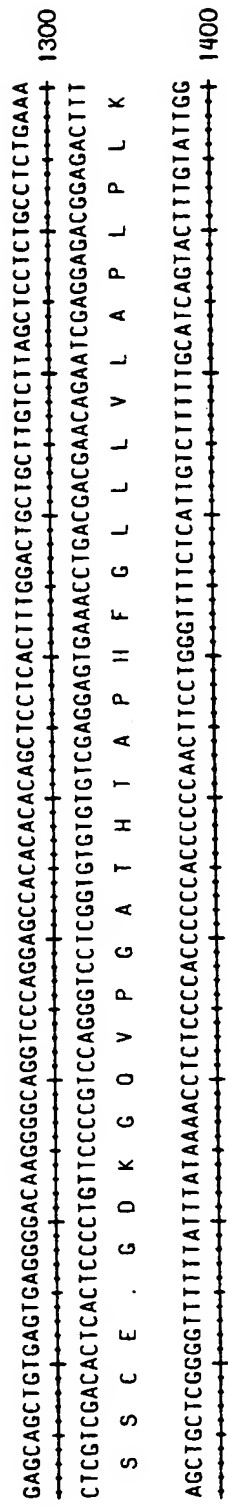
GAGCAGCTGTGAGTGAGGGGACAAGGGCAGGTCCCAGGAGCCACACACACAGCTCCTCACTTTGGACTGCTGCTTGTCTTAGCTCCCTCTGCCTCTGAAA
1300
CTCGTCGACACTCACTCCCCTGTTCCCCGTCCAGGGTCCTCGGTGTGTGTCGAGGAGTGAAACCTGACGACGAACACAGAATCGAGGAGACGGAGACTTT
S S C E . G D K G Q V P G A T H T A P H F G L L L V L A P L P L K
AGCTGCTCGGGGTTTTTTATTTATAAAACCTCTCCCCACCCCCCCAACTTCCTGGGTTTTCTCATTGTCTTTTTGCATCAGTACTTTGTATTGG
1400
TCGACGAGCCCCAAAAAATAAATATTTTGGAGAGGGGTGGGGGGGTTGAAGGACCCAAAAGAGTAACAGAAAAACGTAGTCATGAAACATAACC
S C S G F F I Y K T S P H P P P P N F L G F L I V F L H Q Y F V L
GATATTAAAGAGATTTAACTT
1421
CTATAATTTCTCTAAATTGAA
G Y . R D L T

FIG. 2A-3

```
ATGGCGGCGCTGGGCGGGGACGGGCTGCGTTTACTGTCGGTATCGCGGCCAGAGCGGCAGCCCGAGTCAACCGCGCTGAG
TACCGCCGCGACCCGCCCCTGCCCGACGCAAATGACAGCCATAGCGCCGGTCTCGCCGTCGGGCTCAGTTGGCGCGACTC

 M  A  A  L  G  G  D  G  L  R  L  L  S  V  S  R  P  E  R  Q  P  E  S  T  A  L  S

CGGCCCGGGCTCGGGGCTGTGCTGCTGGGTGTCTGTGTTCTCCTGCTTCAGCCTCGCCTGCTCCTACGTGGGCAGCCTCT
GCCGGGCCCGAGCCCCGACACGACGACCCACAGACACAAGAGGACGAAGTCGGAGCGGACGAGGATGCACCCGTCGGAGA

  G  P  G  S  G  L  C  C  W  V  S  V  F  S  C  F  S  L  A  C  S  Y  V  G  S  L

ACGTGTGGAAGAGCGAGCTGCCCAGGGACCACCCCGCTGTTATCAAGCGGCGTTCCACCAGTGTCCTGGTAGTGTCCAGC
TGCACACCTTCTCGCTCGACGGGTCCCTGGTGGGGCGACAATAGTTCGCCGCAAGGTGGTCACAGGACCATCACAGGTCG

Y  V  W  K  S  E  L  P  R  D  H  P  A  V  I  K  R  R  S  T  S  V  L  V  V  S  S

TTGTCCCCTCTTTGCGTGCTGCTCTGGAGGGAACTCACTGGCATCCAGCCAGGCACATCACTGCTTACCTTGATGGGCTT
AACAGGGGAGAAACGCACGACGAGACCTCCCTTGAGTGACCGTAGGTCGGTCCGTGTAGTGACGAATGGAACTACCCGAA

 L  S  P  L  C  V  L  L  W  R  E  L  T  G  I  Q  P  G  T  S  L  L  T  L  M  G  F

CAGGCTGGAGGGCATTTTCCCAGCAGCGCTGCTGCCCCTGCTGCTAACTATGATCCTTTTCCTGGGTCCACTGATGCAGC
GTCCGACCTCCCGTAAAAGGGTCGTCGCGACGACGGGGACGACGATTGATACTAGGAAAAGGACCCAGGTGACTACGTCG

  R  L  E  G  I  F  P  A  A  L  L  P  L  L  L  T  M  I  L  F  L  G  P  L  M  Q

TCTCTATGGATTGCCCTTGTGACCTGACAGATGGGCTGAAGGTTGTCCTGGCCCCTCGTTCTTGGGCCCGCTGCCTCACA
AGAGATACCTAACGGGAACACTGGACTGTCTACCCGACTTCCAACAGGACCGGGGAGCAAGAACCCGGGCGACGGAGTGT

 L  S  M  D  C  P  C  D  L  T  D  G  L  K  V  V  L  A  P  R  S  W  A  R  C  L  T
```

FIG. 2B-1

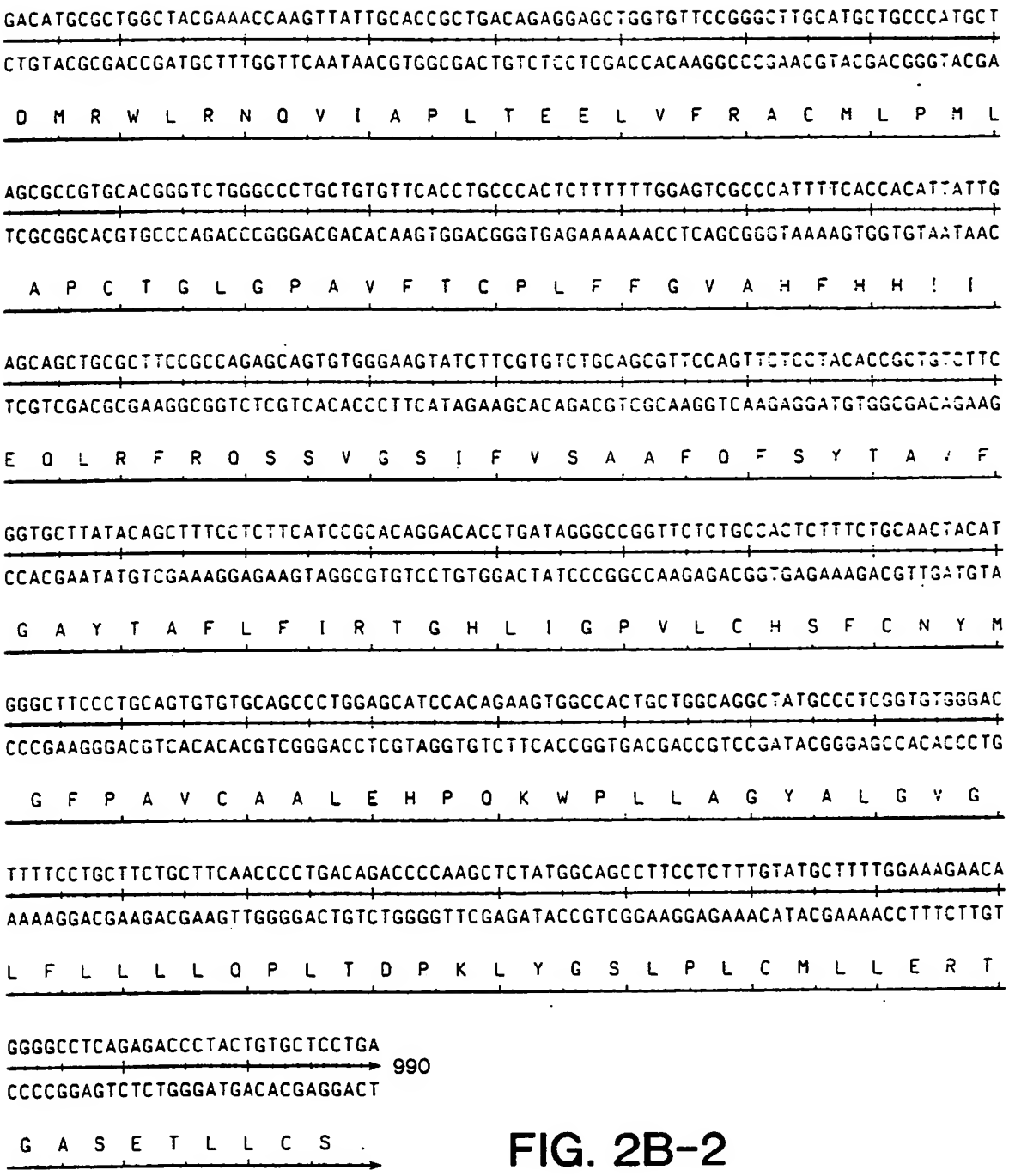
GACATGCGCTGGCTACGAAACCAAGTTATTGCACCGCTGACAGAGGAGCTGGTGTTCCGGGCTTGCATGCTGCCCATGCT
CTGTACGCGACCGATGCTTTGGTTCAATAACGTGGCGACTGTCTCCTCGACCACAAGGCCCGAACGTACGACGGGTACGA
D M R W L R N Q V I A P L T E E L V F R A C M L P M L
AGCGCCGTGCACGGGTCTGGGCCCTGCTGTGTTCACCTGCCCACTCTTTTTTGGAGTCGCCCATTTTCACCACATTATTG
TCGCGGCACGTGCCCAGACCCGGGACGACACAAGTGGACGGGTGAGAAAAAACCTCAGCGGGTAAAAGTGGTGTAATAAC
A P C T G L G P A V F T C P L F F G V A H F H H I I
AGCAGCTGCGCTTCCGCCAGAGCAGTGTGGGAAGTATCTTCGTGTCTGCAGCGTTCCAGTTCTCCTACACCGCTGTCTTC
TCGTCGACGCGAAGGCGGTCTCGTCACACCCTTCATAGAAGCACAGACGTCGCAAGGTCAAGAGGATGTGGCGACAGAAG
E Q L R F R Q S S V G S I F V S A A F Q F S Y T A I F
GGTGCTTATACAGCTTTCCTCTTCATCCGCACAGGACACCTGATAGGGCCGGTTCTCTGCCACTCTTTCTGCAACTACAT
CCACGAATATGTCGAAAGGAGAAGTAGGCGTGTCCTGTGGACTATCCCGGCCAAGAGACGGTGAGAAAGACGTTGATGTA
G A Y T A F L F I R T G H L I G P V L C H S F C N Y M
GGGCTTCCCTGCAGTGTGTGCAGCCCTGGAGCATCCACAGAAGTGGCCACTGCTGGCAGGCTATGCCCTCGGTGTGGGAC
CCCGAAGGGACGTCACACACGTCGGGACCTCGTAGGTGTCTTCACCGGTGACGACCGTCCGATACGGGAGCCACACCCTG
G F P A V C A A L E H P Q K W P L L A G Y A L G V G
TTTTCCTGCTTCTGCTTCAACCCCTGACAGACCCCAAGCTCTATGGCAGCCTTCCTCTTTGTATGCTTTTGGAAAGAACA
AAAAGGACGAAGACGAAGTTGGGGACTGTCTGGGGTTCGAGATACCGTCGGAAGGAGAAACATACGAAAACCTTTCTTGT
L F L L L L Q P L T D P K L Y G S L P L C M L L E R T
GGGGCCTCAGAGACCCTACTGTGCTCCTGA
990
CCCCGGAGTCTCTGGGATGACACGAGGACT
G A S E T L L C S .

FIG. 2B-2

Contig 1:

Contig Length:	1833 bases
Average Length/Sequence:	874 bases
Total Sequence Length:	2622 bases
Top Strand:	1 sequences
Bottom Strand:	2 sequences
Total:	3 sequences

FIG. 2C

```
        10        20        30        40        50        60        70
TTCTGTATTTCAATAAAGCCGTTCCAAGCACAAAGTATCAACTATCGTGCAAGGGGAGGGGCGCCCGGGG 70
GACCCGACAACTCGCGCCGCCGCCTGACGTCACCTACGCCGTCGCGTCACTGCCGCGCGCCACGGGTCCG 140
GGCGTGATGGCGGCGCTGGGCGGGGACGGGCTGCGTTTACTGTCGGTATCGCGGCCAGAGCGGCAGCCCG 210
AGTCAACCGCGCTGAGCGGCCCGGGCTCGGGGCTGTGCTGCTGGGTGTCTGTGTTCTCCTGCTTCAGCCT 280
CGCCTGCTCCTACGTGGGCAGCCTCTACGTGTGGAAGAGCGAGCTGCCCAGGGACCACCCCGCTGTTATC 350
        360       370       380       390       400       410       420
AAGCGGCGTTCCACCAGTGTCCTGGTAGTGTCCAGCTTGTCCCCTCTTTGCGTGCTGCTCTGGAGGGAAC 420
TCACTGGCATCCAGCCAGGCACATCACTGCTTACCTTGATGGGCTTCAGGCTGGAGGGCATTTTCCCAGC 490
AGCGCTGCTGCCCCTGCTGCTAACTATGATCCTTTTCCTGGGTCCACTGATGCAGCTCTCTATGGATTGC 560
CCTTGTGACCTGACAGATGGGCTGAAGGTTGTCCTGGCCCCTCGTTCTTGGGCCCGCTGCCTCACAGACA 630
TGCGCTGGCTACGAAACCAAGTTATTGCACCGCTGACAGAGGAGCTGGTGTTCCGGGCTTGCATGCTGCC 700
        710       720       730       740       750       760       770
CATGCTAGCGCCGTGCACGGGTCTGGGCCCTGCTGTGTTCACCTGCCCACTCTTTTTTGGAGTCGCCCAT 770
TTTCACCACATTATTGAGCAGCTGCGCTTCCGCCAGAGCAGTGTGGGAAGTATCTTCGTGTCTGCAGCGT 840
TCCAGTTCTCCTACACCGCTGTCTTCGGTGCTTATACAGCTTTCCTCTTCATCCGCACAGGACACCTGAT 910
AGGGCCGGTTCTCTGCCACTCTTTCTGCAACTACATGGGCTTCCCTGCAGTGTGTGCAGCCCTGGAGCAT 980
CCACAGAAGTGGCCACTGCTGGCAGGCTATGCCCTCGGTGTGGGACTTTTCCTGCTTCTGCTTCAACCCC 1050
        1060      1070      1080      1090      1100      1110      1120
TGACAGACCCCAAGCTCTATGGCAGCCTTCCTCTTTGTATGCTTTTGGAAAGAACAGGGGCCTCAGAGAC 1120
CCTACTGTGCTCCTGA 1136
```

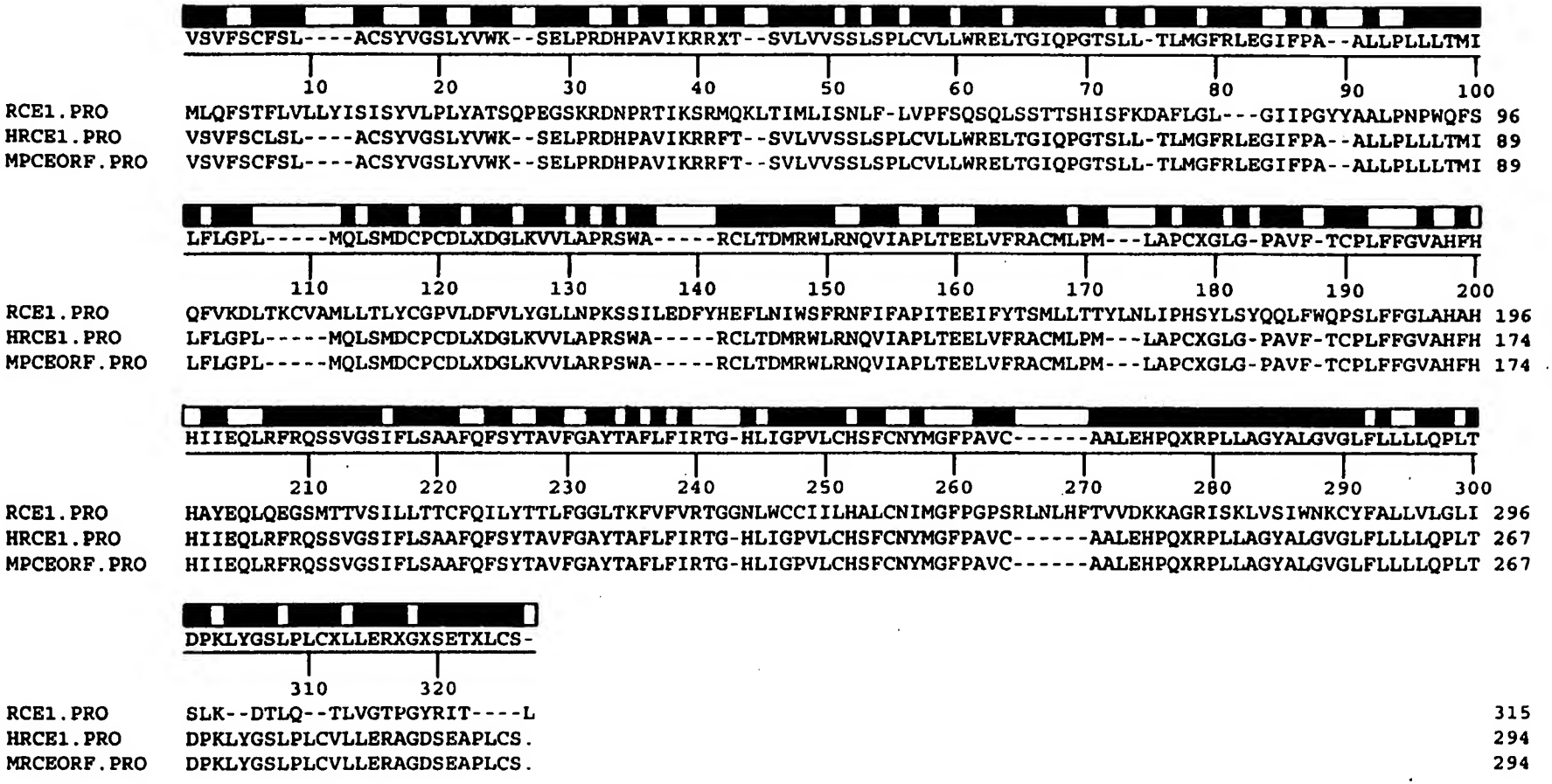
VSVFSCFSL----ACSYVGSLYVWK--SELPRDHPAVIKRRXT--SVLVVSSLSPLCVLLWRELTGIQPGTSLL-TLMGFRLEGIFPA--ALLPLLLTMI
10 20 30 40 50 60 70 80 90 100
RCE1.PRO MLQFSTFLVLLYISISYVLPLYATSQPEGSKRDNPRTIKSRMQKLTIMLISNLF-LVPFSQSQLSSTTSHISFKDAFLGL---GIIPGYYAALPNPWQFS 96
HRCE1.PRO VSVFSCLSL----ACSYVGSLYVWK--SELPRDHPAVIKRRFT--SVLVVSSLSPLCVLLWRELTGIQPGTSLL-TLMGFRLEGIFPA--ALLPLLLTMI 89
MPCEORF.PRO VSVFSCFSL----ACSYVGSLYVWK--SELPRDHPAVIKRRFT--SVLVVSSLSPLCVLLWRELTGIQPGTSLL-TLMGFRLEGIFPA--ALLPLLLTMI 89
LFLGPL-----MQLSMDCPCDLXDGLKVVLAPRSWA-----RCLTDMRWLRNQVIAPLTEELVFRACMLPM---LAPCXGLG-PAVF-TCPLFFGVAHFH
110 120 130 140 150 160 170 180 190 200
RCE1.PRO QFVKDLTKCVAMLLTLYCGPVLDFVLYGLLNPKSSILEDFYHEFLNIWSFRNFIFAPITEEIFYTSMLLTTYLNLIPHSYLSYQQLFWQPSLFFGLAHAH 196
HRCE1.PRO LFLGPL-----MQLSMDCPCDLXDGLKVVLAPRSWA-----RCLTDMRWLRNQVIAPLTEELVFRACMLPM---LAPCXGLG-PAVF-TCPLFFGVAHFH 174
MPCEORF.PRO LFLGPL-----MQLSMDCPCDLXDGLKVVLARPSWA-----RCLTDMRWLRNQVIAPLTEELVFRACMLPM---LAPCXGLG-PAVF-TCPLFFGVAHFH 174
HIIEQLRFRQSSVGSIFLSAAFQFSYTAVFGAYTAFLFIRTG-HLIGPVLCHSFCNYMGFPAVC------AALEHPQXRPLLAGYALGVGLFLLLLQPLT
210 220 230 240 250 260 270 280 290 300
RCE1.PRO HAYEQLQEGSMTTVSILLTTCFQILYTTLFGGLTKFVFVRTGGNLWCCIILHALCNIMGFPGPSRLNLHFTVVDKKAGRISKLVSIWNKCYFALLVLGLI 296
HRCE1.PRO HIIEQLRFRQSSVGSIFLSAAFQFSYTAVFGAYTAFLFIRTG-HLIGPVLCHSFCNYMGFPAVC------AALEHPQXRPLLAGYALGVGLFLLLLQPLT 267
MPCEORF.PRO HIIEQLRFRQSSVGSIFLSAAFQFSYTAVFGAYTAFLFIRTG-HLIGPVLCHSFCNYMGFPAVC------AALEHPQXRPLLAGYALGVGLFLLLLQPLT 267
DPKLYGSLPLCXLLERXGXSETXLCS-
310 320
RCE1.PRO SLK--DTLQ--TLVGTPGYRIT----L 315
HRCE1.PRO DPKLYGSLPLCVLLERAGDSEAPLCS. 294
MRCEORF.PRO DPKLYGSLPLCVLLERAGDSEAPLCS. 294

FIG. 3

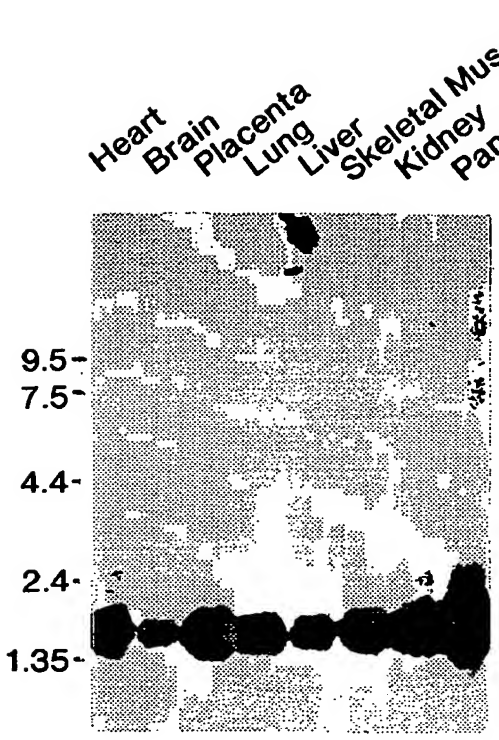
Heart
Brain
Placenta
Lung
Liver
Skeletal Muscle
Kidney
Pancreas
9.5
7.5
4.4
2.4
1.35

FIG. 4A

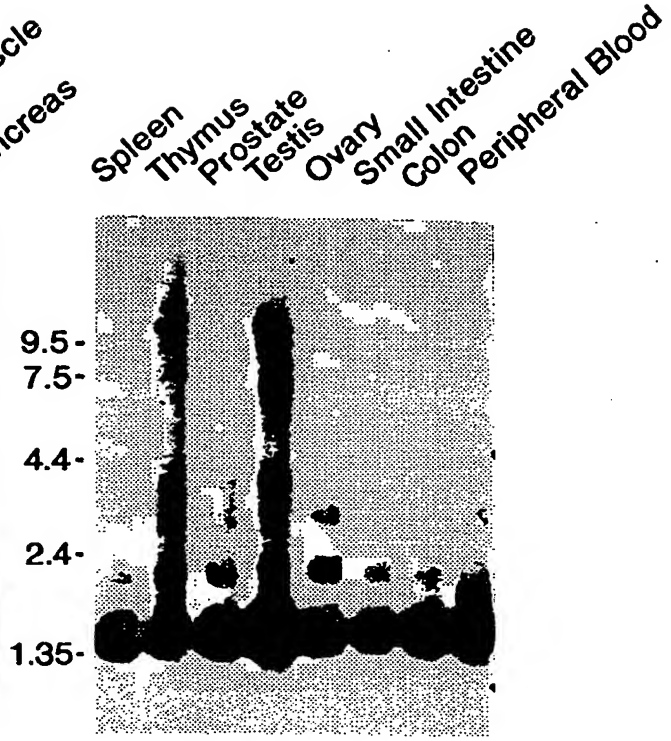
Spleen
Thymus
Prostate
Testis
Ovary
Small Intestine
Colon
Peripheral Blood
9.5
7.5
4.4
2.4
1.35

FIG. 4B

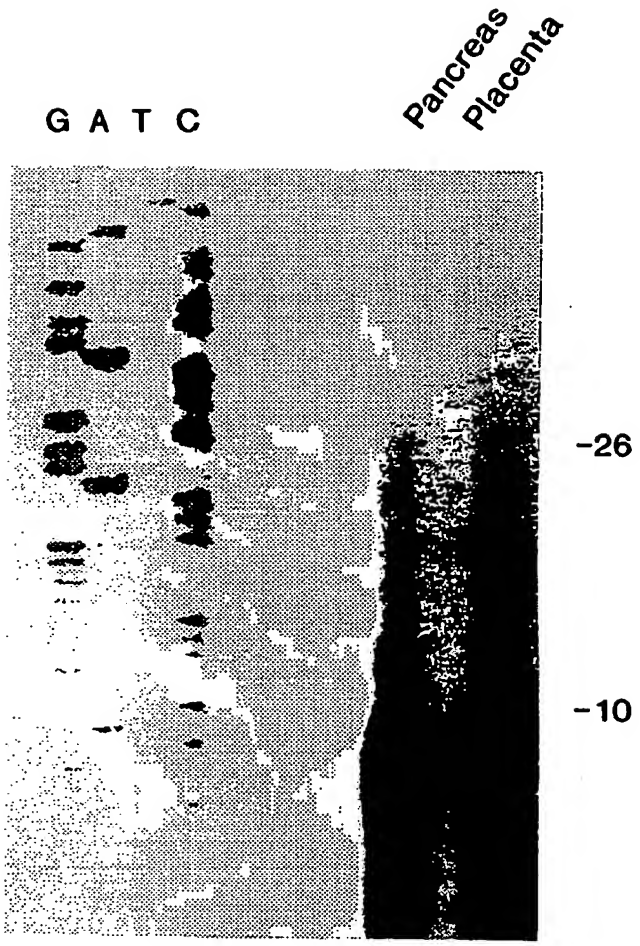

FIG. 5

W303

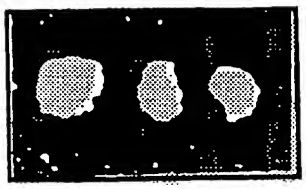

FIG. 6A

ΔΔ

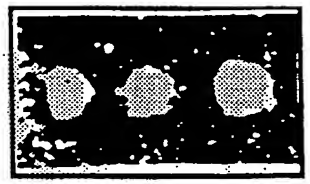

FIG. 6B

Δ *afc1*

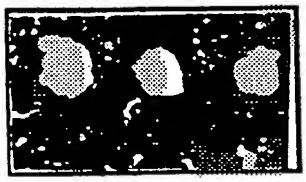

FIG. 6C

Δ *rce1*

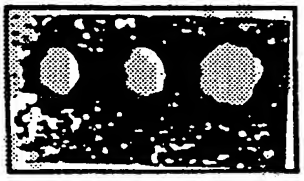

FIG. 6D

ΔΔ pJP1

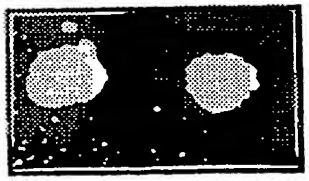

FIG. 6E

ΔΔpACA7001

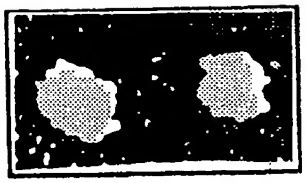

FIG. 6F

ΔΔpACA7002

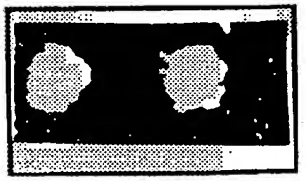

FIG. 6G

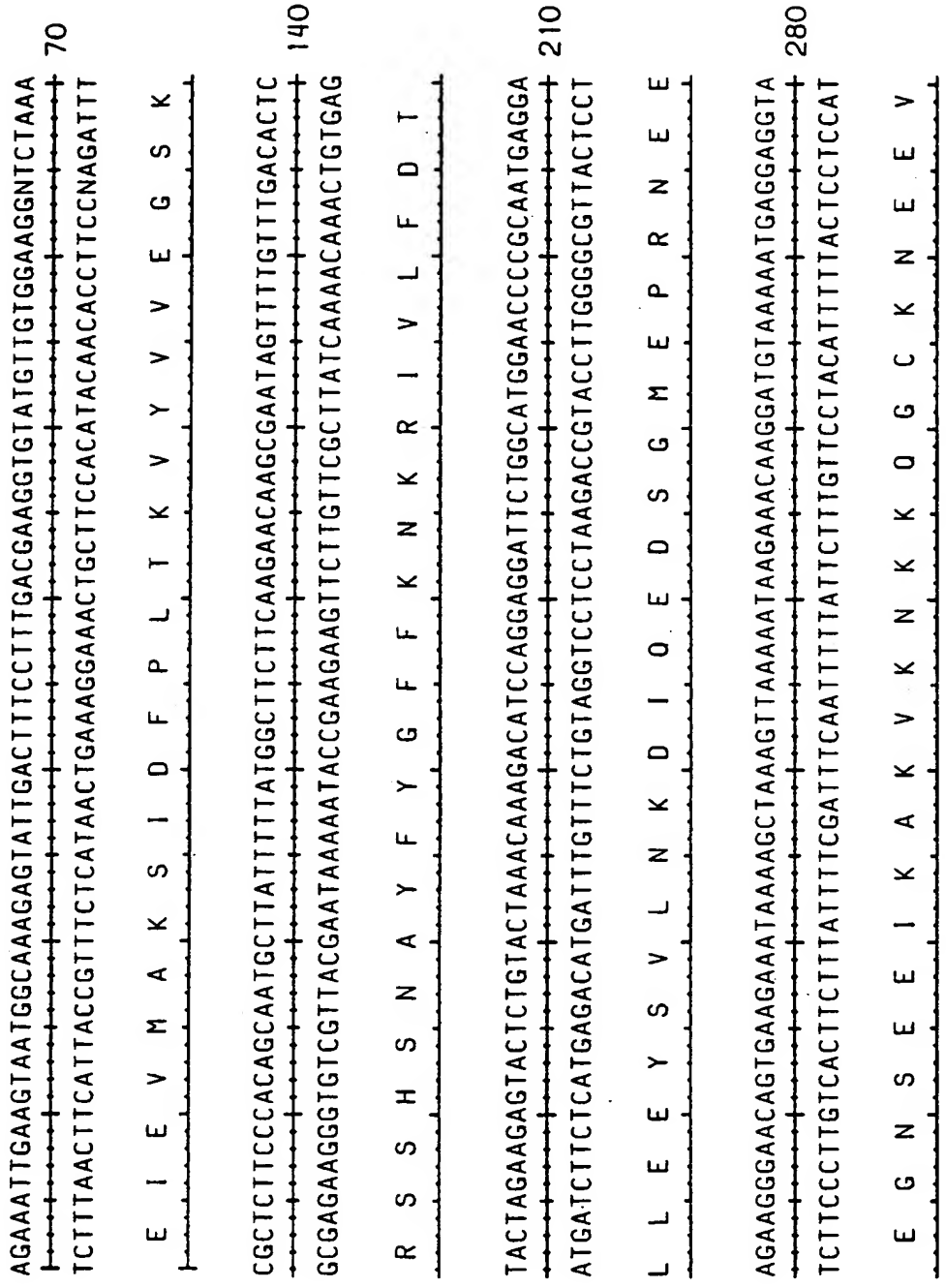
AGAAATTGAAGTAATGGCAAAGAGTATTGACTTTCCTTTGACGAAGGTGTATGTTGTGGAAGGNTCTAAA
TCTTTAACTTCATTACCGTTTCTCATAACTGAAAGGAAACTGCTTCCACATACAACACCTTCCNAGATTT
70
E I E V M A K S I D F P L T K V Y V V E G S K
CGCTCTTCCCACAGCAATGCTTATTTTTATGGCTTCTTCAAGAACAAGCGAATAGTTTTGTTTGACACTC
GCGAGAAGGGTGTCGTTACGAATAAAAATACCGAAGAAGTTCTTGTTCGCTTATCAAAACAAACTGTGAG
140
R S S H S N A Y F Y G G F F K N K R I V L F D T
TACTAGAAGAGTACTCTGTACTAAACAAAGACATCCAGGAGGATTCTGGCATGGAACCCCGCAATGAGGA
ATGATCTTCTCATGAGACATGATTTGTTTCTGTAGGTCCTCCTAAGACCGTACCTTGGGGCGTTACTCCT
210
L L E E Y S V L N K D I O E D S G M E P R N E E
AGAAGGGAACAGTGAAGAAATAAAAGCTAAAGTTAAAAATAAGAAACAAGGATGTAAAAATGAGGAGGTA
TCTTCCCTTGTCACTTCTTTATTTTCGATTTCAATTTTTATTCTTTGTTCCTACATTTTTACTCCTCCAT
280
E G N S E E I K A K V K N K K O G C K N E E V

FIG. 7A

```
CTCGCTGTACTAGGCCATGAACTGGGGCACTGGAAGTTGGGACATACAGTCAAAAATATCATTATTAGCC
GAGCGACATGATCCGGTACTTGACCCCGTGACCTTCAACCCTGTATGTCAGTTTTTATAGTAATAATCGG
                                                                      350

  L  A  V  L  G  H  E  L  G  H  W  K  L  G  H  T  V  K  N  I  I  I  S
  ------------------------------------------------------------------

AGATGAATTCTTTCCTGTGTTTTTTTTATTTGCTGTATTAATTGGTCGAAAGGAGCTTTTTGCTGCATT
TCTACTTAAGAAAGGACACAAAAAAAATAAACGACATAATTAACCAGCTTTCCTCGAAAAACGACGTAA
                                                                      420

  O  M  N  S  F  L  L  C  F  F  L  F  A  V  L  I  G  R  K  E  L  F  A  A  F
  ------------------------------------------------------------------

TGGTTTTTATGATAGCCAACCCACTCTTATTGGACTATTGATCATCTTCCAGTTTATTTTTTCACCTTAC
ACCAAAAATACTATCGGTTGGGTGAGAATAACCTGATAACTAGTAGAAGGTCAAATAAAAAAGTGGAATG
                                                                      490

  G  F  Y  D  S  O  P  T  L  I  G  L  L  I  I  F  O  F  I  F  S  P  Y
  ------------------------------------------------------------------

AATGAGGTTCTTTCTTTTTGCCTAACAGTCCTAAGCCGCAGATTTGAGTTTCAAGCTGATGCATTTGCCA
TTACTCCAAGAAAGAAAAACGGATTGTCAGGATTCGGCGTCTAAACTCAAAGTTCGACTACGTAAACGGT
                                                                      560

  N  E  V  L  S  F  C  L  T  V  L  S  R  R  F  E  F  O  A  D  A  F  A
  ------------------------------------------------------------------
```

FIG. 7B

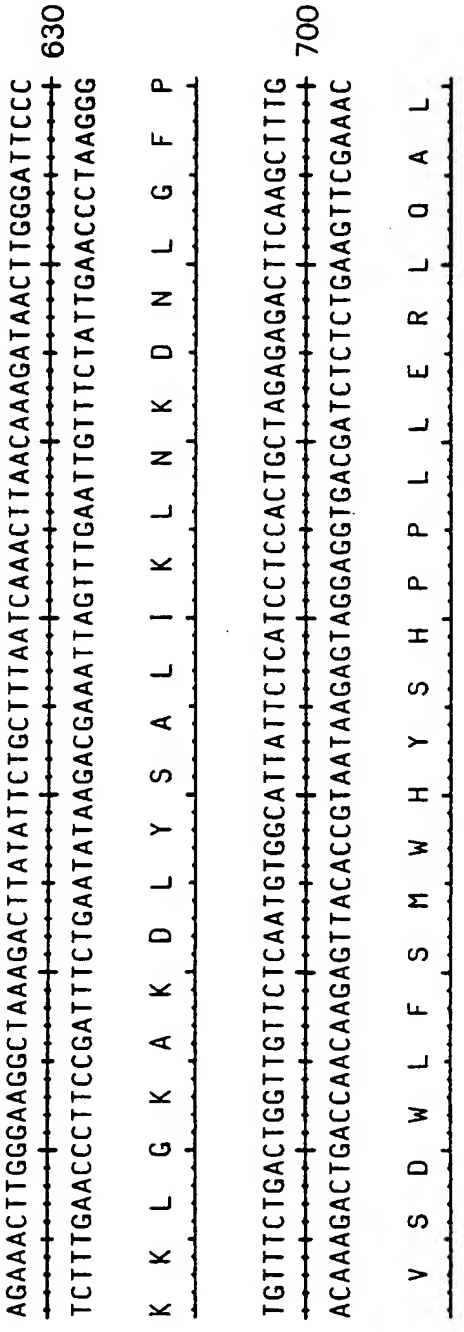
AGAAACTTGGGAAGGCTAAAGACTTATATTCTGCTTTAATCAAACTTAACAAAGATAACTTGGGATTCCC
TCTTTGAACCCTTCCGATTTCTGAATATAAGACGAAATTAGTTTGAATTGTTTCTATTGAACCCTAAGGG
630
K K L G K A K D L Y S A L I K L N K D N L G F P
TGTTTCTGACTGGTTGTTCTCAATGTGGCATTATTCTCATCCTCCACTGCTAGAGAGACTTCAAGCTTTG
ACAAAGACTGACCAACAAGAGTTACACCGTAATAAGAGTAGGAGGTGACGATCTCTCTGAAGTTCGAAAC
700
V S D W L F S M W H Y S H P P L L E R L Q A L
AAAACTATGAAGCAACACTGA
TTTTGATACTTCGTTGTGACT
721
K T M K Q H .

FIG. 7C

FIG. 8

SUBSTITUTE SHEET (RULE 26)

```
                                                                  Score     E
Sequences producing significant alignments:                       (bits)  Value

sp|Q03530|RCE1_YEAST  CAAX PRENYL PROTEASE 2 (PRENYL PROTEIN-SPE...    70  9e-13
sp|P47154|ST24_YEAST  CAAX PRENYL PROTEASE 1 (PRENYL PROTEIN-SPE...    28  4.1


 sp|Q03530|RCE1_YEAST CAAX PRENYL PROTEASE 2 (PRENYL PROTEIN-SPECIFIC ENDOPROTEASE 2)
          (PPSEP 2) (RAS AND A-FACTOR CONVERTING ENZYME) (RACE)
          Length = 315

 Score = 69.9 bits (168), Expect = 9e-13
 Identities = 33/70 (47%), Positives = 45/70 (64%), Gaps = 1/70 (1%)

Query: 6   LFFGVAHFHHIIEQLRFRQSSVGNIFLSAAFQFSYTAVFGAYTAFLFIRT-GHLIGPVLC 64
           LFFG+AH HH  EQL+    +  +I L+  FQ  YT +FG  T F+F+RT G+L   ++
Sbjct: 188 LFFGLAHAHHAYEQLQEGSMTTVSILLTTCFQILYTTLFGGLTKFVFVRTGGNLWCCIIL 247

Query: 65  HSFCNYMGFP 74
           H+ CN MGFP
Sbjct: 248 HALCNIMGFP 257


 sp|P47154|ST24_YEAST CAAX PRENYL PROTEASE 1 (PRENYL PROTEIN-SPECIFIC ENDOPROTEASE 1)
          (PPSEP 1) (A-FACTOR CONVERTING ENZYME)
          Length = 453

 Score = 27.8 bits (60), Expect = 4.1
 Identities = 19/73 (26%), Positives = 31/73 (42%), Gaps = 2/73 (2%)

Query: 3   TCPLFFGVAHF--HHIIEQLRFRQSSVGNIFLSAAFQFSYTAVFGAYTAFLFIRTGHLIG 60
           T  L   + H+  +HI+  + F Q      IF       +  T+ +  +  FL   TG  +
Sbjct: 292 TAVLAHEIGHWQKNHIVNMVIFSQLHTFLIFSLFTSIYRNTSFYNTFGFFLEKSTGSFVD 351

Query: 61  PVLCHSFCNYMGF 73
           PV+   F   +GF
Sbjct: 352 PVITKEFPIIIGF 364
```

FIG. 9

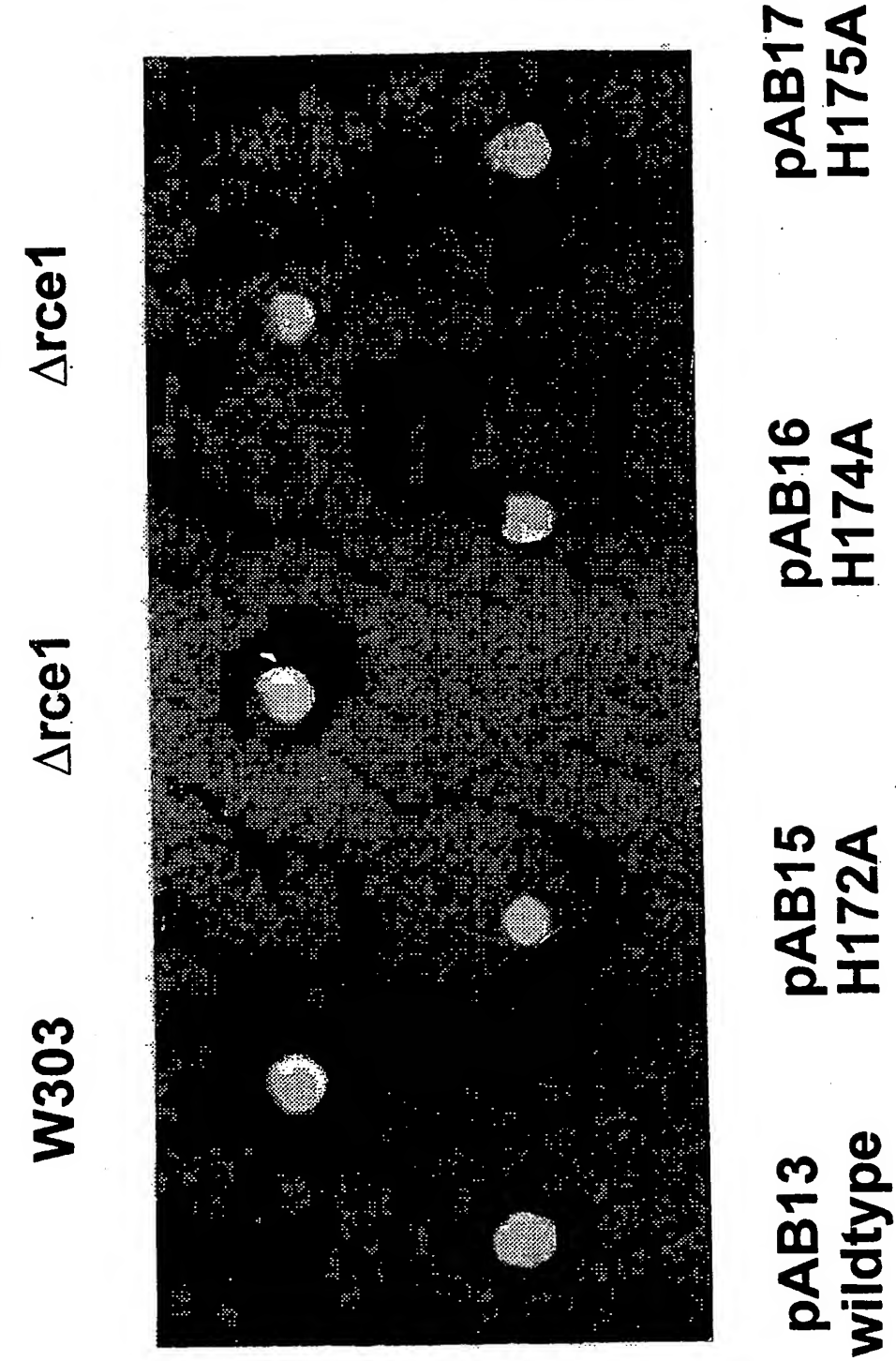
W303
Δrce1
Δafc1
Δrce1
pAB13
wildtype
pAB15
H172A
pAB16
H174A
pAB17
H175A

FIG. 10

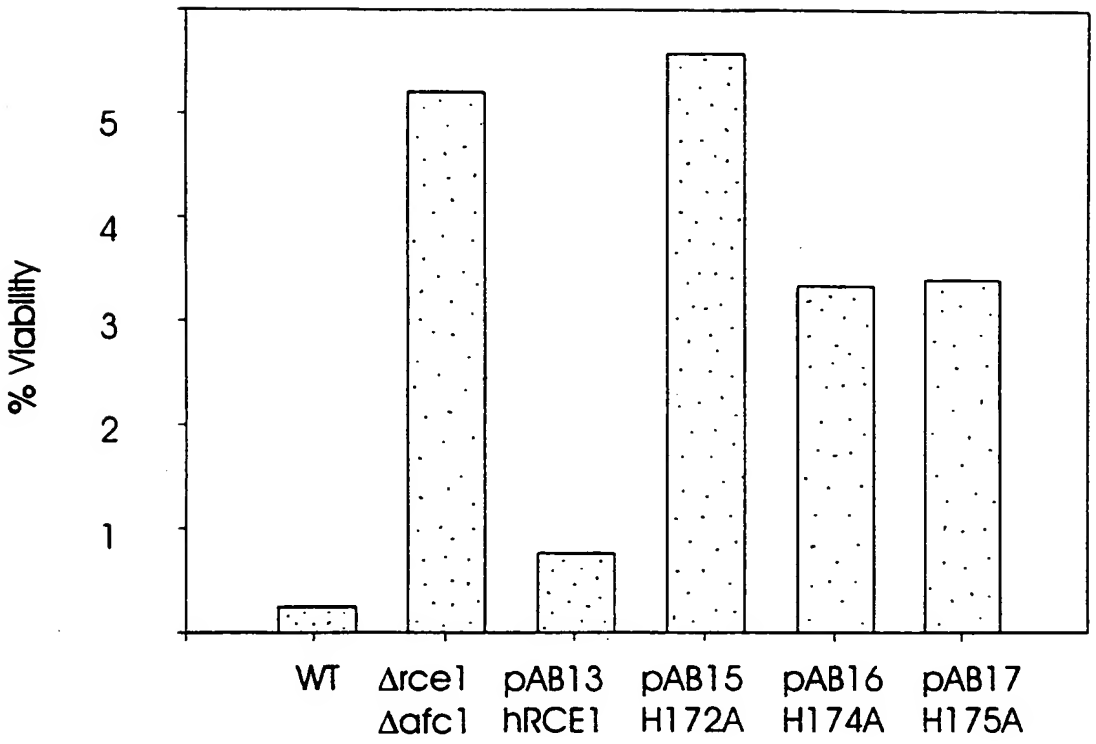
% Viability
5
4
3
2
1
WT
Δrce1 Δafc1
pAB13 hRCE1
pAB15 H172A
pAB16 H174A
pAB17 H175A

FIG. 11

Applicant's or agent's file reference number	ACA2 CIP PCT	International application	PCT/US98/11415

INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13*bis*)

A. The indications made below relate to the microorganism referred to in the description on page 23, lines 4-11

B. IDENTIFICATION OF DEPOSIT Further deposits are identified on an additional sheet [X]

Name of depositary institution

American Type Culture Collection

Address of depositary institution *(including postal code and country)*

10801 University Boulevard
Manassas, Virginia 20110
United States of America

Identification Ref. by Depositor: hRCE1 - isolated from a human total 9 week fetus cDNA library in lambda gt II (ATCC#77433), 27R - λgt II

Date of deposit	Accession Number
7 May 1997 (07.05.97)	209020

C. ADDITIONAL INDICATIONS *(leave blank if not applicable)* This information is continued on an additional sheet [ ]

It is the applicant's wish that, until either a patent has issued on the basis of the application or the application is refused, or is abandoned and no longer subject to reinstatement, or is withdrawn, the Commissioner only authorize the furnishing of a sample of any deposited biological material referred to in the specification to an independent expert nominated by the Commissioner in accordance with section 109 or 165 of the Patent Rules (Rule 104(4) or 160(4)).

D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE *(if the indications are not for all designated States)*

CA

E. SEPARATE FURNISHING OF INDICATIONS *(leave blank if not applicable)*

The indications listed below will be submitted to the International Bureau later *(specify the general nature of the indications e.g., "Accession Number of Deposit")*

For receiving Office use only	For International Bureau use only
[ ] This sheet was received with the international application	[✓] This sheet was received by the International Bureau on: 29 SEP 1998
Authorized officer	Authorized officer Nicola Wolff

Form PCT/RO/134 (July 1992)

Applicant's or agent's file reference number	ACA2 CIP PCT	International application no.	PCT/US98/11415

INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13*bis*)

A. The indications made below relate to the microorganism referred to in the description on page 26 lines 7-13

B. IDENTIFICATION OF DEPOSIT Further deposits are identified on an additional sheet ☐

Name of depositary institution

American Type Culture Collection

Address of depositary institution (*including postal code and country*)

10801 University Boulevard
Manassas, Virginia 20110
United States of America

Identification Ref. by Depositor: pAB13-plasmid in Escherichia coli strain DH5alpha, AB13

Date of deposit	Accession Number
4 June 1998 (04.06.98)	209940

C. ADDITIONAL INDICATIONS (*leave blank if not applicable*) This information is continued on an additional sheet ☐

It is the applicant's wish that, until either a patent has issued on the basis of the application or the application is refused, or is abandoned and no longer subject to reinstatement, or is withdrawn, the Commissioner only authorize the furnishing of a sample of any deposited biological material referred to in the specification to an independent expert nominated by the Commissioner in accordance with section 109 or 165 of the Patent Rules (Rule 104(4) or 160(4)).

D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (*if the indications are not for all designated States*)

CA

E. SEPARATE FURNISHING OF INDICATIONS (*leave blank if not applicable*)

The indications listed below will be submitted to the International Bureau later (*specify the general nature of the indications e.g., "Accession Number of Deposit"*)

For receiving Office use only	For International Bureau use only
☐ This sheet was received with the international application	☑ This sheet was received by the International Bureau on: 29 SEP 1998
Authorized officer	Authorized officer NICOLA WOLFF

Form PCT/RO/134 (July 1992)

Applicant's or agent's file reference number	ACA2 CIP PCT	International application	PCT/US98/11415

INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13*bis*)

A. The indications made below relate to the microorganism referred to in the description on page 23, lines 4-11

B. IDENTIFICATION OF DEPOSIT Further deposits are identified on an additional sheet [X]

Name of depositary institution

American Type Culture Collection

Address of depositary institution *(including postal code and country)*

10801 University Boulevard
Manassas, Virginia 20110
United States of America

Identification Ref. by Depositor: hRCE1 - isolated from a human total 9 week fetus cDNA library in lambda gt II (ATCC#77433), 27R - λgt II

Date of deposit	Accession Number
7 May 1997 (07.05.97)	209020

C. ADDITIONAL INDICATIONS *(leave blank if not applicable)* This information is continued on an additional sheet [ ]

In respect of the designation of the EPO, samples of the deposited microorganisms will be made available until the publication of the mention of the grant of the European patent or until the date on which the application is refused or withdrawn or is deemed to be withdrawn, as provided in Rule 28(3) of the Implementing Regulations under the EPC only by the issue of a sample to an expert nominated by requester (Rule 28(4) EPC).

D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE *(if the indications are not for all designated States)*

EP

E. SEPARATE FURNISHING OF INDICATIONS *(leave blank if not applicable)*

The indications listed below will be submitted to the International Bureau later *(specify the general nature of the indications e.g., "Accession Number of Deposit")*

For receiving Office use only	For International Bureau use only
[ ] This sheet was received with the international application	[X] This sheet was received by the International Bureau on: 29 SEP 1998
Authorized officer	Authorized officer NICOLA J. WOLFF

Form PCT/RO/134 (July 1992)

Applicant's or agent's file reference number	ACA2 CIP PCT	International application	PCT/US98/11415

INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13*bis*)

A. The indications made below relate to the microorganism referred to in the description on page 26, lines 7-13

B. IDENTIFICATION OF DEPOSIT Further deposits are identified on an additional sheet ☐

Name of depositary institution

American Type Culture Collection

Address of depositary institution *(including postal code and country)*

10801 University Boulevard
Manassas, Virginia 20110
United States of America

Identification Ref. by Depositor: pAB13-plasmid in Escherichia coli strain DH5alpha, AB13

Date of deposit	Accession Number
4 June 1998 (04.06.98)	209940

C. ADDITIONAL INDICATIONS *(leave blank if not applicable)* This information is continued on an additional sheet ☐

In respect of the designation of the EPO, samples of the deposited microorganisms will be made available until the publication of the mention of the grant of the European patent or until the date on which the application is refused or withdrawn or is deemed to be withdrawn, as provided in Rule 28(3) of the Implementing Regulations under the EPC only by the issue of a sample to an expert nominated by requester (Rule 28(4) EPC).

D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE *(if the indications are not for all designated States)*

EP

E. SEPARATE FURNISHING OF INDICATIONS *(leave blank if not applicable)*

The indications listed below will be submitted to the International Bureau later *(specify the general nature of the indications e.g., "Accession Number of Deposit")*

For receiving Office use only	For International Bureau use only
☐ This sheet was received with the international application	☑ This sheet was received by the International Bureau on: 29 SEP 1998
Authorized officer	Authorized officer NICOLA J. WOLFF

Form PCT/RO/134 (July 1992)

Applicant's or agent's file reference number	ACA2 CIP PCT	International application	PCT/US98/11415

INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13*bis*)

A. The indications made below relate to the microorganism referred to in the description on page 23, lines 4-11

B. IDENTIFICATION OF DEPOSIT — Further deposits are identified on an additional sheet [X]

Name of depositary institution

American Type Culture Collection

Address of depositary institution (*including postal code and country*)

10801 University Boulevard
Manassas, Virginia 20110
United States of America

Identification Ref. by Depositor: hRCE1 - isolated from a human total 9 week fetus cDNA library in lambda gt II (ATCC#77433), 27R - λgt II

Date of deposit	Accession Number
7 May 1997 (07.05.97)	209020

C. ADDITIONAL INDICATIONS (*leave blank if not applicable*) — This information is continued on an additional sheet [ ]

Applicant(s) hereby give notice of my/our intention that samples of the above-identified culture shall be available only to experts in accordance with paragraph 3 of the Fourth Schedule to the Patents Rules 1995.

D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (*if the indications are not for all designated States*)

SG

E. SEPARATE FURNISHING OF INDICATIONS (*leave blank if not applicable*)

The indications listed below will be submitted to the International Bureau later (*specify the general nature of the indications e.g., "Accession Number of Deposit"*)

For receiving Office use only	For International Bureau use only
[ ] This sheet was received with the international application	[✓] This sheet was received by the International Bureau on: 29 SEP 1998
Authorized officer	Authorized officer NICOLA WOLFF

Form PCT/RO/134 (July 1992)

Applicant's or agent's file reference number	A_A2 CIP PCT	International application	PCT/US98/11415

INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13*bis*)

A. The indications made below relate to the microorganism referred to in the description on page 26, lines 7-13

B. IDENTIFICATION OF DEPOSIT — Further deposits are identified on an additional sheet ☐

Name of depositary institution

American Type Culture Collection

Address of depositary institution (*including postal code and country*)

10801 University Boulevard
Manassas, Virginia 20110
United States of America

Identification Ref. by Depositor: pAB13-plasmid in Escherichia coli strain DH5alpha, AB13

Date of deposit	Accession Number
4 June 1998 (04.06.98)	209940

C. ADDITIONAL INDICATIONS (*leave blank if not applicable*) — This information is continued on an additional sheet ☐

Applicant(s) hereby give notice of my/our intention that samples of the above-identified culture shall be available only to experts in accordance with paragraph 3 of the Fourth Schedule to the Patents Rules 1995.

D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (*if the indications are not for all designated States*)

SG

E. SEPARATE FURNISHING OF INDICATIONS (*leave blank if not applicable*)

The indications listed below will be submitted to the International Bureau later (*specify the general nature of the indications e.g., "Accession Number of Deposit"*)

For receiving Office use only	For International Bureau use only
☐ This sheet was received with the international application	☑ This sheet was received by the International Bureau on: 23 SEP 1998
Authorized officer	Authorized officer Nicola Wolff

Form PCT/RO/134 (July 1992)

PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : C12N 15/57, 9/64, C07K 16/40, C12Q 1/37, A61K 35/00, 38/48 — A3

(11) International Publication Number: WO 98/54333

(43) International Publication Date: 3 December 1998 (03.12.98)

(21) International Application Number: PCT/US98/11415

(22) International Filing Date: 2 June 1998 (02.06.98)

(30) Priority Data:

60/047,369	2 June 1997 (02.06.97)	US
60/052,389	14 July 1997 (14.07.97)	US

(63) Related by Continuation (CON) or Continuation-in-Part (CIP) to Earlier Applications

US	60/047,369 (CIP)
Filed on	2 June 1997 (02.06.97)
US	60/052,389 (CIP)
Filed on	14 July 1997 (14.07.97)

(71) Applicant *(for all designated States except US)*: ACACIA BIOSCIENCES INC. [US/US]; 4136 Lakeside Drive, Richmond, CA 94806 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: ASHBY, Matthew, N. [US/US]; 91 Longfellow Road, Mill Valley, CA 94941 (US). DIMSTER-DENK, Dago, G. [US/US]; 904 Sir Francis Drake Boulevard #8, Kentfield, CA 94904 (US). PHILIPS, John, W. [US/US]; 10 Marinita Avenue, San Rafael, CA 94901 (US).

(74) Agents: HALEY, James, F., Jr. et al.; Fish & Neave, 1251 Avenue of the Americas, New York, NY 10020 (US).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, GW, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

Published

With international search report.

Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

With an indication in relation to a deposited biological material furnished under Rule 13bis separately from the description.

(88) Date of publication of the international search report: 8 April 1999 (08.04.99)

(54) Title: MAMMALIAN CAAX PROCESSING ENZYMES

(57) Abstract

The present invention provides mammalian DNA sequences that display a high degree of sequence identity to their *Saccharomyces cerevisae* counterparts, *RCE1 and AFC1*. Specifically, complementary DNA (cDNA) sequences of the human and mouse *RCE* functional homologs are provide. Human cDNA sequences encoding proteins having a high degree of amino acid sequence identity to the yeast Afc1p protein are also provided. This invention is also directed to recombinant DNA molecules comprising the mammalian DNA sequences, DNA molecules and antisense RNA molecules which hybridize under stringent hybridization conditions to those DNA sequences, hosts transformed with their recombinant DNA molecules and protein expression products produced by culturing the transformed hosts. Antibodies directed against the protein expression products are also provided. Also provided are assays to identify inhibitors of one or more mammalian CAAX processing enzymes and kits for making the above products and performing the above assays. Finally, this invention provides pharmaceutical compositions comprising an inhibitor of a mammalian CAAX processing enzyme, and methods for treating a CAAX-protein mediated disease or disorder in a patient by administering such a pharmaceutical composition..

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	**ES**	Spain	**LS**	Lesotho	**SI**	Slovenia
AM	Armenia	**FI**	Finland	**LT**	Lithuania	**SK**	Slovakia
AT	Austria	**FR**	France	**LU**	Luxembourg	**SN**	Senegal
AU	Australia	**GA**	Gabon	**LV**	Latvia	**SZ**	Swaziland
AZ	Azerbaijan	**GB**	United Kingdom	**MC**	Monaco	**TD**	Chad
BA	Bosnia and Herzegovina	**GE**	Georgia	**MD**	Republic of Moldova	**TG**	Togo
BB	Barbados	**GH**	Ghana	**MG**	Madagascar	**TJ**	Tajikistan
BE	Belgium	**GN**	Guinea	**MK**	The former Yugoslav Republic of Macedonia	**TM**	Turkmenistan
BF	Burkina Faso	**GR**	Greece			**TR**	Turkey
BG	Bulgaria	**HU**	Hungary	**ML**	Mali	**TT**	Trinidad and Tobago
BJ	Benin	**IE**	Ireland	**MN**	Mongolia	**UA**	Ukraine
BR	Brazil	**IL**	Israel	**MR**	Mauritania	**UG**	Uganda
BY	Belarus	**IS**	Iceland	**MW**	Malawi	**US**	United States of America
CA	Canada	**IT**	Italy	**MX**	Mexico	**UZ**	Uzbekistan
CF	Central African Republic	**JP**	Japan	**NE**	Niger	**VN**	Viet Nam
CG	Congo	**KE**	Kenya	**NL**	Netherlands	**YU**	Yugoslavia
CH	Switzerland	**KG**	Kyrgyzstan	**NO**	Norway	**ZW**	Zimbabwe
CI	Côte d'Ivoire	**KP**	Democratic People's Republic of Korea	**NZ**	New Zealand		
CM	Cameroon			**PL**	Poland		
CN	China	**KR**	Republic of Korea	**PT**	Portugal		
CU	Cuba	**KZ**	Kazakstan	**RO**	Romania		
CZ	Czech Republic	**LC**	Saint Lucia	**RU**	Russian Federation		
DE	Germany	**LI**	Liechtenstein	**SD**	Sudan		
DK	Denmark	**LK**	Sri Lanka	**SE**	Sweden		
EE	Estonia	**LR**	Liberia	**SG**	Singapore		

INTERNATIONAL SEARCH REPORT

International Application No
PCT/US 98/11415

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 C12N15/57 C12N9/64 C07K16/40 C12Q1/37 A61K35/00
A61K38/48

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 6 C12N C07K C12Q A61K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	Database EMBL Emest16, Entry MMAA68614 Accession number AA168614 22 December 1996 90% identity with Seq.ID:1 nt.390-839 100% identity with Seq.ID:3 nt.383-832 XP002083133 cited in the application	1,3-17
Y	see the whole document ---	20
X	Database EMBL Emest13, Entry HSW4115 Accession number W96411 18 July 1996 98% identity with Seq.ID:1 nt.488-841 88% identity with Seq.ID:3 nt.481-830 reverse orientation XP002083134 cited in the application	1-17
Y	see the whole document ---	20
	-/--	

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
1 February 1999	23.02.99

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Macchia, G

2

Form PCT/ISA/210 (second sheet) (July 1992)

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	Database EMBL Emest15, Entry MM3448 Accession number W14344 30 April 1996 87% identity with Seq.ID:1 nt.617-990 100% identity with Seq.ID:3 nt.610-983 XP002083135 cited in the application	1,3-17
Y	see the whole document	20

X	CHEN Y. ET AL.: "Solubilization, partial purification, and affinity labeling of the membrane-bound isoprenylated protein endoprotease" BIOCHEMISTRY, vol. 35, 12 March 1996, pages 3227-3237, XP000199958	16,19,22
Y	see page 3227	20

X	AKOPYAN T.N. ET AL.: "Proteolytic processing of farnesylated peptides: assay and partial purification from pig brain membranes of an endopeptidase which has the characteristics of E.C. 3.4.24.15" BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS, vol. 198, no. 2, 28 January 1994, pages 787-794, XP002049582 cited in the application	16,19,22
Y	see abstract	20

X	NISHII W. ET AL.: "Partial purification and characterization of a CAAX-motif-specific protease from bovine brain using a novel fluorometric assay" JOURNAL OF BIOCHEMISTRY, vol. 122, 1997, pages 402-408, XP000199959	16,19,22
Y	see the whole document	20

X	Database EMBL Emest12, Entry HSAA19855 Accession number AA210930 2 February 1997 96% identity with Seq.ID:37 nt.216-590 XP002091645	1,3-17
Y	see the whole document	20

X	Database EMBL Emest8, Entry HS181310 Accession number N76181 5 April 1996 100% identity with Seq.ID:37 nt.430-721 XP002091646	1,3-17
Y	see the whole document	20

.	-/--	

2

C.(Continuation) DOCUMENTS CONSIDERED T BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	Database EMBL Emest8, Entry HS1228356 Accession number AA428315 25 May 1997 100% identity with Seq.ID:1 nt.130-529 90% identity with Seq.ID:3 nt.123-522 XP002082655 see the whole document ---	1-6,8-11
X	EP 0 679 716 A (MATSUBARA KENICHI; OKUBO KOUSAKU (JP)) 2 November 1995 Seq.ID:985 is 96% identical to Seq.ID:1 nt.1090-1421 see abstract see page 1 - page 9 ---	1-6,8-11
X	Database EMBL Emest16, Entry MMAA21859 Accession number AA021859 29 November 1996 89% identity with Seq.ID:1 nt.7-664 98% identity with Seq.ID:3 nt.3-672 XP002083136 see the whole document ---	,
X	Database EMBL Emest18, Entry HSZZ64182 Accession number AA359058 18 April 1997 100% identity with Seq.ID:37 nt.1-306 XP002091647 see the whole document ---	1,3-6, 8-11
E	EP 0 887 415 A (SMITHKLINE BEECHAM CORP SMITHKLINE BEECHAM PLC (US) KIKLY KNAB SOUTHAN) 30 December 1998 see the whole document ---	1-22
E	EP 0 887 416 A (SMITHKLINE BEECHAM CORP SMITHKLINE BEECHAM PLC (US) KIKLY KNAB SOUTHAN) 30 December 1998 see the whole document ---	1,3-22
P,X	WO 98 05786 A (UNIVERSITY OF CALIFORNIA; RINE J.D.; BOYARTCHUK V.L.; ASHBY M.N. (US)) 12 February 1998 see page 12, line 21 - page 13, line 22 see page 20, line 9 - page 22, line 32 see page 34, line 21 - page 37, line 22 see page 54 - page 57; claims --- -/--	1,3-17, 19,20,22

2

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	BOYARTCHUK V.L. ET AL.: "Modulation of Ras and a-factor function by carboxyl-terminal proteolysis" SCIENCE, vol. 275, 21 March 1997, pages 1796-1800, XP002049583 cited in the application see page 1799, right-hand column, paragraph 4 - page 1800, left-hand column, paragraph 1 -----	1,19,20, 22

INTERNATIONAL SEARCH REPORT

In. .ational application No.
PCT/US 98/11415

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. [X] Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

Remark: although claims 21 and 22 are directed to a method of treatment of the human/animal body, the search has been carried out and based on the alleged effects of the compound/composition.

2. [ ] Claims Nos.:
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:

3. [ ] Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

see additional sheet

1. [X] As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.

2. [ ] As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. [ ] As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:

4. [ ] No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

[ ] The additional search fees were accompanied by the applicant's protest.

[X] No protest accompanied the payment of additional search fees.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

1. Claims: 2 totally; 1, 3-22 all partially.

   DNA molecule comprising a sequence as in Seq.ID:1, 3 or 5, homologous or hybridizing sequences thereof. Vector comprising said DNA molecule, host cell transformed with said vector. Polypeptide encoded by said DNA molecules, comprising aminoacid sequence as in Seq.ID:2, 4 or 6 or homologous to said Seq.IDs; corresponding method of production. Method for identifying a competitive inhibitor of said polypeptide. Pharmaceutical composition comprising said polypeptide, or an inhibitor, application in therapy thereof. Antibody specific for an epitope of said polypeptide.

2. Claims: 1, 3-22 all partially.

   As invention 1 but concerning Seq.ID:37 and 38.

INTERNATIONAL SEARCH REPORT

mation on patent family members

International Application No
PCT/US 98/11415

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
EP 0679716 A	02-11-1995	AU 8116494 A	13-06-1995
		CA 2153480 A	01-06-1995
		WO 9514772 A	01-06-1995
EP 0887415 A	30-12-1998	NONE	
EP 0887416 A	30-12-1998	NONE	
WO 9805786 A	12-02-1998	AU 4157297 A	25-02-1998